Stochastic Multi-Market Modeling with 'Efficient Quadratures': Does the Rotation of Stroud's Octahedron Matter?

DISSERTATION

zur Erlangung des akademischen Grades
doctor rerum agriculturarum (Dr. rer. agr.)

eingereicht an der
Landwirtschaftlich-Gärtnerischen Fakultät
der Humboldt-Universität zu Berlin
von

Marco Antonio Artavia Oreamuno (MSc)

Präsident der Humboldt-Universität zu Berlin
Prof. Dr. Jan-Hendrik Olbertz

Dekan der Landwirtschaftlich-Gärtnerischen Fakultät
Prof. Dr. Dr. h.c. Frank Ellmer

Gutachter: 1. Prof. Dr. Dr. h.c. Dieter Kirschke
2. Prof. Dr. Harald Grethe
3. Dr. Martin Banse

Tag der mündlichen Prüfung: 27.02.2012

Preface

'Efficient quadratures' and numerical integration of multivariate functions in the European Simulation Model (ESIM) has not been an easy topic but a very interesting one.

I would like to express my gratitude to my supervisor Prof. Dr. Dr. h. c. Dieter Kirschke for his incredible support and supervision, both in the subject specific area and in everyday life at the Department of Agricultural Policy. Also, I would like to thank Prof. Dr. Harald Grethe for encouraging me to conduct different projects related with this topic, as well as for his dedicated supervision and the productive exchange of ideas.

In addition, I would like to thank Prof. Dr. Georg Zimmermann for his mathematical support, which made it possible to slowly work through the topic.

Very special thanks go to Thordis Möller and Uli Kleinwechter for their support and friendship.

Also, very special thanks go to Kerstin Oertel who was always ready to help and give advice on technical problems.

Additionally, my thanks go to Till Below and Lukas Scholz for helping me with the logistics associated with the final phase of the dissertation.

I would like to express my gratitude to my family and friends Hadaluz, Gabriel, Mariajosé, Annabelle, Sami, Aida, Micha, Jörg and Florian for their patience and for being there when I needed them.

Finally, many thanks go to my wife Kristina and my children Lukas, Camilo and Lucía for going with me on this long way of ups and downs, for their patience and unconditional support.

Table of contents

List of Abbreviations

CV	Coefficient of variation
ESIM	European simulation model
EU	European Union
EV	Expected value
LHS	Latin Hypercube Sampling
MCS	Monte Carlo Sampling
NMS	The EU New Member States covered in the enlargements of 2004 and 2007
t	Tonnes
TRQ	Tariff rate quota
PDF	Probability density function
ROW	The 'rest of the world' region in ESIM

Notation

The notation is divided into 4 groups: 1) notation for spatial regions, 2) matrix and vector notation, 3) notation used to describe formulas in ESIM, and 4) other notation.

1) Notation for spatial regions

B^n The unit n-ball – the set of all **x** such that (Haber, 1970):

$$(x_1)^2 + (x_2)^2 + \cdots + (x_n)^2 \leq 1$$

C^n The n-cube with a centroid at the origin and vertices:

$$(\pm a, \pm a, \cdots, \pm a)$$

G^n The unit n-cube – the set of all **x** such that (Haber, 1970):

$$0 \leq x_i \leq 1 \qquad \text{for } i = 1,2, \cdots, n$$

H^n A n-rectangle defined by the Cartesian product of n closed intervals (Trench, 2011):

$$H^n = [a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n]$$

R Any region of integration – analogous to an interval $[a, b]$ for $n = 1$

$|R|$ Measure of the region R of integration – e.g., area and volume for $n = 2$ and $n = 3$ respectively

$\mathbb{R}^n$ Euclidean n-dimensional space

2) Matrix and vector notation

In general,

i) matrices are denoted with upper case, bold letters;

ii) column vectors with lower case, bold letters;

iii) row vectors with lower case, bold, italic letters and a subscript indicating the row or variable; and

ix) vectors of row vectors with lower case, bold, italic letters, but without a subscript.

The following are some specificities:

$\mathbf{I}_n$ The identity matrix of size $n \times n$

i Index to determine the elements (variables or coordinates) of a column vector:

$$i = 1,2,\cdots,n$$

k Index to determine the quadrature points in a matrix of quadratures – these are column vectors:

$$k = 1,2,\cdots,N$$

L Lower triangular matrix of the Cholesky factorization: $\boldsymbol{\Sigma}[\boldsymbol{z}] = \mathbf{L}\mathbf{L}^{\mathrm{T}}$

N Number of quadrature points or required evaluations of an integrand

n Number of variables in a column vector - the number of dimensions of a multivariate integration problem

P[ ] Correlation matrix of the data in the content of the brackets

R Upper triangular matrix of the reverse Cholesky factorization:

$$\boldsymbol{\Sigma}[\boldsymbol{z}] = \mathbf{R}\mathbf{R}^{\mathrm{T}}$$

$[\,]^{\mathrm{T}}$ Transpose

U The orthogonal matrix consisting of the eigenvectors of $\boldsymbol{\Sigma}[\boldsymbol{z}]$ – these are situated in the columns of **U**

X Matrix of quadrature points for the approximation of a multivariate normal PDF:

$$\begin{bmatrix} x_{1,1} & \cdots & x_{1,N} \\ \vdots & \ddots & \vdots \\ x_{n,1} & \cdots & x_{n,N} \end{bmatrix}$$

Z Matrix of deviates with n variables and m observations:

$$\begin{bmatrix} z_{1,1} & \cdots & z_{1,m} \\ \vdots & \ddots & \vdots \\ z_{n,1} & \cdots & z_{n,m} \end{bmatrix}$$

Γ Matrix of quadrature points for the multivariate standard normal PDF or the C^n with vertices $(\pm 1, \pm 1, \cdots, \pm 1)$:

$$\begin{bmatrix} \gamma_{1,1} & \cdots & \gamma_{1,N} \\ \vdots & \ddots & \vdots \\ \gamma_{n,1} & \cdots & \gamma_{n,N} \end{bmatrix}$$

μ The vector of means of the stochastic variables in a multivariate PDF:

$$\boldsymbol{\mu} = (\mu_1, \mu_2, \cdots, \mu_n)$$

Σ[ ] The covariance matrix of the data in the content of the brackets:

$$\begin{bmatrix} \sigma_{1,1} & \cdots & \sigma_{1,n} \\ \vdots & \ddots & \vdots \\ \sigma_{n,1} & \cdots & \sigma_{n,n} \end{bmatrix}$$

where, $\sigma_{i,i} = \sigma_i^2 = \mathrm{Var}(x_i)$

0 The zero vector with n-elements:

$$\mathbf{0} = (0,0,\cdots,0)^{\mathrm{T}}$$

3) Notation used to described formulas in ESIM

$elastsp_{crops,crops}$	Own and cross price elasticity of supply
$elastyd$	Own price elasticity of yield
$elastyi$	Yield elasticity with respect to the index of intermediate input costs
$elastyl$	Yield elasticity with respect to the index of labor costs
int. ind	Cost index of intermediated inputs
lab. ind	Cost index of labor inputs
NX	Net exports
PD	Domestic price
P_lo	Lower price level of the price logistic function (the largest value between intervention price or world market price)
PP	Producer price
P_up	Upper price level of the price logistic function for products with tariffs (the largest value between threshold price or world market price)
P_up2	Price level given by the sum of world market price and export subsidy

$subsquant$	Export subsidy limit
$\mathrm{sup.int}$	Constant in the supply function (it stands for supply intercept)
$\mathrm{tp.gr}$	Technical progress of yields or supply
$TRADSHR$	Trade share
TRQ	Tariff Rate Quota
$TUSE$	Total use
$\mathrm{yild.int}$	Constant in the yield function (it stands for yield intercept)
α	Parameter in the logistic price function to control the steepness of the slope between P_{up} and P_{lo}
β	Parameter in the logistic price function to control the symmetry/asymmetry of the slope between P_{up} and P_{lo}

4) Other notation

d	Degree of precision of a quadrature formula
E[ ]	Expected value of the data in the content of the brackets
Q^n	The n-octahedron from Stroud's theorem from 1957
Q_t^n	The transformed Q^n for the approximation of the multivariate normal distributions, obtained through: $Q_t^n = \mathbf{A\Gamma}$
x	An independent variable
$\mathbf{x}$	A vector of variables whose components are denoted by subscripts: $\mathbf{x} = (x_1, x_2, \cdots, x_n)$
$f(x)$	The univariate PDF
$f(\mathbf{x})$	The multivariate PDF of the random vector $\mathbf{x}$

List of Tables

List of Figures

List of Boxes

Abstract

Recently, stochastic applications of large-scale applied simulation models of agricultural markets have become more common. However, stochastic modeling with large market models incurs high computational and management costs for data storage, analysis and manipulation. Gaussian Quadratures (GQ) are efficient sampling methods requiring few points to approximate the central moments of the joint probability distribution of stochastic variables, and therefore reduce computational costs. For symmetric regions of integration, the vertices of Stroud's n-octahedron (Stroud 1957) are formulas of degree 3 with minimal number of points, which can make the stochastic modeling with large economic models manageable. However, the conjecture exists that rotations of Stroud's n-octahedron may have an effect on the accuracy of approximation of the model results. To address this, eight different rotations (quadrature formulas) were tested using the European Simulation Model (ESIM). It was found that using the formulas from Artavia et al. (2009) or Arndt (1996) in the generation of the quadratures is crucial, and furthermore, that the formula from Arndt yields higher accuracy. With the rotation obtained with Arndt's formula and in models or markets with high asymmetries, as is the case for soft wheat in ESIM, the arrangement of the stochastic variables (A1 or A2) in the covariance matrix or the method selected to induce the covariance matrix (via Cholesky decomposition – C - or via the diagonalization method – D -) may have a significant effect on the accuracy of the quadratures. With Arndt's formula and with less asymmetric markets, as is the case for rapeseed in ESIM, the selection of arrangements A1 or A2 and of the method to induce the covariance C or D might not have a significant effect on the accuracy of the quadratures.

Keywords: Sampling Methods, Gaussian Quadratures, Monte Carlo, Stochastic Modeling, Commodity Markets

Kurzfassung

Stochastische Anwendungen von großen Simulationsmodellen des Agrarsektors werden immer häufiger. Allerdings ist die stochastische Modellierung mit großen Marktmodellen rechenintensiv und mit hohen Kosten für Datenabspeicherung, -analyse und -manipulation verbunden. Gausssche Quadraturen sind effiziente Stichprobenmethoden, die wenige Punkte für die Approximation der zentralen Momente von gemeinsamen Wahrscheinlichkeitsverteilungen brauchen und somit die Kosten der Datenmanipulation senken. Für symmetrische Integrationsräume sind die Ecken des Oktaeder von Stroud (Stroud 1957) Formeln dritten Grades mit minimaler Anzahl von Punkten, die die stochastische Modellierung mit großen Modellen handhabbar machen kann. Es gibt trotzdem die Vermutung, dass Rotationen von Stroud's Oktaeder einen Einfluss auf die Exaktheit der Quadraturen haben könnten; daher werden in dieser Studie acht unterschiedliche Rotationen (Quadraturformeln) getestet. Es zeigte sich, dass der Gebrauch der Formel von Artavia et al. (2009) oder der von Arndt (1996) bei der Generierung der Quadraturen entscheidend ist, und dass die Formel von Arndt einen höheren Exaktheitsgrad ergibt. Mit der Rotation, die sich aus der Formel von Arndt ergibt und Modellen oder Märkten mit starken Asymmetrien wie der Weizenmarkt in ESIM, könnten die Reihenfolge der stochastischen Variablen in der Kovarianz Matrix (A1 oder A2) oder die Methoden zur Einführung der Kovarianz Matrix (via Cholesky-Zerlegung –C- oder via die Diagonalisierungsmethode –D-) einen bedeutsamen Einfluss auf die Exaktheit der Quadraturen haben. Mit Arndt's Formel und weniger asymmetrischen Modellen oder Märkten, wie der Fall von Raps in ESIM, haben die Reihenfolgen A1 und A2 oder die Methoden zur Einführung der Kovarianz Matrix C und D weniger Einfluss auf die Exaktheit der Quadraturen.

Schlagwörter: Stichprobenmethoden, Gausssche Quadraturen, Monte Carlo, stochastische Modellierung, Agrarmärkte

1 Introduction

1.1 Presentation of the problem

Since 2007, a significant increase in market instability has been observed and this has highlighted the importance of including uncertainty in the different analyses carried out with multi-market simulation models. The actual economic crises of the industrial world, as well as the observed price and weather extremes in the last decade, have awoken the interest of politicians and researchers in studying this uncertainty and its effects. For example, some large economic models that include the agricultural markets – such as the European Simulation Model (ESIM), the Aglink-Cosimo model, the FAPRI partial equilibrium model and GTAP – have started to incorporate stochastic features in order to consider market uncertainty. Referring specifically to ESIM, this model has been used in different research projects concerned with stochastic modeling. The European Commission has financed a project in order to develop a stochastic version of it (see Artavia et al., 2008), and it has been used for the analysis of a possible increase in extreme weather conditions and their effect on price volatility. The last, in the context of a study by the Edmund-Rehwinkel Foundation on the consequences of the financial crisis and unstable markets on agriculture (see Artavia et al., 2010b).

The Aglink-Cosimo model, in its version presented in OECD (2003), and the FAPRI partial equilibrium model (FAPRI-MU, 2011) use Monte Carlo methods for the approximation of multivariate PDFs of stochastic variables, while ESIM and GTAP (Hertel et al., 2005; and Hertel et al., 2010) use Gaussian Quadratures in their stochastic versions. Note that in this study, these are also called 'efficient quadratures'.

'Efficient quadratures' present the advantage that they approximate the moments of the multivariate PDFs with significantly fewer points. Thus, these are of special interest for the numerical evaluation of integrals of functions of multivariate random variables with high n, which may be posed by stochastic formulations of large economic models. For example, if in such models it is intended to capture some of the relevant sources of the variability of agricultural markets, then stochastic variables can be attached to yield, production area and input prices on the supply side, and to income and population growth on the

demand side. If these data are gathered for every agricultural commodity and for every country or region in the model, one rapidly deals with integration problems of high dimensionality. In these cases, it is desirable to keep the number of evaluations of the integrand, N, as small as possible – to save on time and computational requirements. As a result, 'efficient quadratures' are an important approach which can make some large integration problems more manageable.

Stroud (1957) proposed to use the vertices of an n-octahedron, Q^n, as the quadrature points of an integration formula of degree 3 of precision[1] for symmetrical regions. This formula uses the least possible number of points, $N = 2n$ (Mysovskih, 1966, as cited by Haber, 1970) and it currently offers the simplest means for conducting stochastic modeling with 'efficient quadratures' in large models (Arndt, 1996). Nonetheless, it presents the disadvantage that it's theoretical foundations, as well as those of the 'efficient quadratures' in general, are rather complex. This problem results in relatively few large models with stochastic versions – for models such as AGMEMOD, CAPSIM, CAPRI and IMPACT[2], no stochastic versions have been published (Artavia et al., 2009) – as well as in relatively few models and applications making use of Stroud's Q^n or other 'efficient quadratures'.

There are different formulas to get Stroud's Q^n. For example, Arndt (1996) and Artavia et al. (2009)[3] apply two different quadrature formulas for the approximation of the multivariate standard normal PDF. Also, Artavia et al. (2009) present different formulas to induce a desired correlation to the quadratures (via Cholesky decomposition of the covariance matrix, via the diagonalization method, and via the reverse Cholesky decomposition), which result in transformed n-octahedrons of different shape and rotation. Moreover, while working with Artavia et al.'s formula in combination with the Cholesky decomposition (in Artavia et al., 2009; Artavia et al., 2010a; and Artavia et al., 2010b), it was realized that the variables, i, from $i = 1$ to $i = n$, are approximated with a progressively increasing diversity of values (values different from zero). As a result, the following question emerged: 'do different arrangements of the stochastic variables in the matrix of deviates produces different results of the multi-market model?'. The same question also emerged

[1] Note that for convenience in some cases the 'degree of precision' of a formula will be simply referred as the 'degree' of the formula.

[2] See Verhoog et al. (2008) for references on the models.

[3] To improve the readability, the formulas from Arndt (1996) and Artavia et al. (2009) will be written without their year of publication.

for the different approaches taken to generate the quadratures: 'are different quadrature formulas equal in terms of the accuracy of model results?'. These questions have not yet been analyzed, even though the Q_t^n obtained with different formulas is already used in combination with multi-market models (ESIM and GTAP).

Another important topic related to stochastic modeling is the level of correlation of the stochastic variables and, particularly, how to induce it to the selected quadratures. This subject has not been treated clearly in the literature (Artavia et al., 2009; Preckel et al., 2010; Horridge and Pearson, 2011); also, the how to do it" has not yet been spelled in detail (Horridge and Pearson, 2011).

1.2 Objectives of the Study

Primary objective

To determine whether different formulas to obtain the transformed *n*-octahedron from Stroud (different quadratures of degree 3 of precision) have consequences on the accuracy of approximation of the results of a multi-market, partial equilibrium model of the agricultural sector.

Secondary objectives

i) To summarize the theoretical background on the numerical integration of single and multiple integrals.

ii) To carry out a theoretical analysis of the formulas from Arndt and Artavia et al. These formulas generate samples of the multivariate standard normal PDF and are used as basis points for the generation of the quadratures for the approximation of specific multivariate normal PDFs.

iii) To clarify misleading instructions found in the literature of stochastic modeling on how to induce a desired covariance matrix to the quadratures.

iv) To document the stochastic module from the ESIM 2011 version. For the analysis of the accuracy of the quadratures tested, the stochastic version of ESIM is used. This version has been continuously developed in different projects (Artavia et al., 2008; Artavia et al., 2009; Artavia et al., 2010a; and Artavia et al., 2010b), but not all the development steps have been described.

v) To illustrate with an example the relevance and gain of information from conducting uncertainty analyses with multi-market models.

1.3 Structure of the Study

Chapter 2 presents a theoretical review of numerical integration, which provides the background necessary for understanding 'efficient quadratures'. In it, the main numerical integration approaches are presented, with a view to outlining their characteristics and applicability.

Chapter 3 conducts a theoretical analysis of the quadrature formulas from Arndt and Artavia et al. This chapter also covers how to induce the desired correlation to the quadratures. By presenting different methods, it clarifies the current misleading instructions given in the literature and using an example it shows the "how to do it" in detail.

Chapter 4 describes the analytical tool used for the study of the main research question; it documents and explains the relevant topics around the stochastic formulation of ESIM.

Chapters 5 and 6 both present simulation analyses with ESIM.

Chapter 5 investigates the accuracy of different quadrature formulas. First, the true values of the results of the multi-market model are estimated using LHS methods. The estimated true values are then used for the determination of the approximation error of the quadrature formulas tested. The results are presented and discussed.

Chapter 6 presents an application of the stochastic version of ESIM. It gives an example of what it is possible to do with stochastic multi-market models. The chapter analyses the consequences of the liberalization of the EU cereals regime on price levels and price uncertainty.

Chapter 7 summarizes the main conclusions and briefly describes the future research agenda.

Clarification of Cooperation Work

Please note that the following work was done in cooperation with other researchers.

The expositions on how to induce a desired covariance matrix to the quadratures of Section 3.2 are based on work done together with Thordis Möller (Ph.D. student at Humboldt University of Berlin), Prof. Dr. Harald Grethe (Prof. of agricultural economics at the University of Hohenheim and the main developer of the current version of ESIM), and Prof. Dr. Georg Zimmermann (Prof. of mathematics at the University of Hohenheim, who contributed strongly to the understanding of quadratures and the methods to induce correlation); see Artavia et al., 2009. Sections 4.1 and 4.3 are descriptions based on the work prepared in

the context of the project "Including Stochastic Elements in the Model Analysis", prepared for the European Commission (see Artavia et al., 2008). Sections 4.2 and 4.4 are based on the work done in the context of the project for the Edmund-Rehwinkel Foundation and the paper prepared for the 50th conference of the German Society of Agricultural Economists (GeWiSoLa) (see Artavia et al., 2010b; and Artavia et al., 2010a, respectively).

2 Theoretical Background on Numerical Integration

In stochastic modeling, when determining the expected value or variance of model results, the researcher is confronted with a numerical integration problem.

For example, let $g(x)$ represent a model function dependent on a random variable, and let $f(x)$ represent its PDF. Then the problem is to find a numerical approximation for the mean (see Arndt, 1996):

$$\mathrm{E}[g(x)] = \int_{-\infty}^{\infty} g(x)f(x)dx, \tag{2.1}$$

and for the variance:

$$\mathrm{E}[(g(x) - \mathrm{E}[g(x)])^2] = \int_{-\infty}^{\infty} (g(x) - \mathrm{E}[g(x)])^2 f(x)\, dx. \tag{2.2}$$

In large multi-market equilibrium models it is very difficult to evaluate (2.1) and (2.2) analytically. The embedded system of model equations, dimensionality and possible implemented constraints to functions will normally make it impossible. Therefore, the approach often used is to develop discrete approximations (samples) of the multivariate PDFs, to evaluate the functions at the generated points and, through weighted sums, to approximate the central moments of the model results. The discrete approximations of the multivariate PDFs are computed using numerical integration methods.

In this chapter, the background necessary for the comprehension of numerical integration formulas is presented. The main approaches are explained to show their characteristics and applicability.

2.1 Numerical Evaluation of Single Integrals

2.1.1 Basic Concepts

Numerical integration is a branch of numerical analysis concerned with giving numerical approximations to definite integration problems:

$$\int_a^b g(x)dx, \tag{2.3}$$

which are not expressed in tractable forms (Hildebrand, 1987). This occurs, e.g., if the function is not well behaved (not continuous or without upper and lower bounds in $[a, b]$) or if a polynomial describing it does not exist (only the function values at some determined evaluation points are known).

All numerical integration methods solve the integration problems using discretization methods which approximate the integral by finite sums corresponding to some partition of the interval of integration $[a, b]$ (Stoer and Bulirsch, 2002). Thus, due to their aim of defining and summing rectangle areas, the formulas obtained with these methods are also called quadratures. Normally, these formulas will give a set of evaluation points, $x_1, x_2, \cdots, x_N$, with associated weights, $w_1, w_2, \cdots, w_N$, so that the integration problem can be approximated by[4]:

$$\int_a^b g(x)dx \approx \sum_{k=1}^{N} w_k g(x_k) \qquad for\ k = 1,2, \cdots, N. \tag{2.4}$$

One basic approach to solve the integration problem is the Riemann sum. In this case, the area bounded by the curve of the function, the x-axis, and the interval $[a, b]$, is subdivided into rectangles with base h_k and height $g(x_k)$. Then, the sum of the area of the rectangles will give an approximation of the area under the curve as follows (Trench, 2011):

$$\int_a^b g(x)dx \approx \sum_{k=1}^{N} h_k g(x_k)\,. \tag{2.5}$$

Note that the evaluation points $x_1, x_2, \cdots, x_N$ can be chosen arbitrarily in $h_1, h_2, \cdots, h_N$, thus, there are infinitely many Riemann sums for the defined partition of $[a, b]$ (Trench, 2011). If we increase the number of rectangles by decreasing the step size for all h, the integral of $g(x)$, if it exists, will be approximated more accurately.

[4] The symbol " ≈" will be used to mean "is approximately equal to".

The Newton-Cotes formulas and the Gaussian quadratures are the basic integration rules and a closer look at their structures will help us to understand how numerical integration algorithms work.

In the case of the Newton-Cotes formulas, equally spaced evaluation points, $x_1, x_2, \cdots, x_N$, in $[a, b]$ are used to obtain $g(x_1), g(x_2), \cdots, g(x_N)$. Then, a polynomial $P(x)$ of degree $N-1$ or lower is interpolated, and whose function values coincide with $g(x)$'s values at the evaluation points, so that $P(x_k) = g(x_k)$ for all k. In this way, a curve which fills in the gaps between the function values $g(x_1), g(x_2), \cdots, g(x_N)$ is interpolated. With this procedure, it is hoped to generate a $P(x)$ such that:

$$P(x) \approx g(x), \tag{2.6}$$

and with it, it is also hoped that (Hildebrand, 1987):

$$\int_a^b g(x)dx \approx \int_a^b P(x)dx. \tag{2.7}$$

The interpolated polynomial $P(x)$ is obtained with the Lagrange interpolation formula (Stoer and Bulirsch, 2002):

$$P(x) = L(x) = \sum_{k=1}^{N} g(x_k) l_k(x), \tag{2.8}$$

where, $L(x)$ is the interpolated polynomial in the Lagrange form and the $l_k(x)$ are the Lagrange basis polynomials.

Moreover, the Lagrange basis polynomials are computed with the formula (Stoer and Bulirsch, 2002):

$$l_k(x) = \prod_{\substack{k=1 \\ k \neq j}}^{N} \frac{x - x_j}{x_k - x_j}. \tag{2.9}$$

For example, supposing a three points x_1, x_2, x_3 Newton-Cotes formula, then, (Stoer and Bulirsch, 2002):

$$l_1(x) = \frac{x - x_2}{x_1 - x_2} \cdot \frac{x - x_3}{x_1 - x_3} \tag{2.10}$$

$$l_2(x) = \frac{x - x_1}{x_2 - x_1} \cdot \frac{x - x_3}{x_2 - x_3}$$

$$l_3(x) = \frac{x - x_1}{x_3 - x_1} \cdot \frac{x - x_2}{x_3 - x_2}.$$

Finally, once $P(x)$ has been computed, it can be integrated, and, as stated in (2.7), it is hoped that its integral in $[a, b]$ is approximately equal to the integral of $g(x)$, also in $[a, b]$.

In (2.11) (own development), it can be observed that, as stated by Stoer and Bulirsch (2002), the weights, $w_1, w_2, \cdots, w_N$, are obtained when computing $\int_a^b l_k(x)dx$ for all k:

$$\begin{aligned} \int_a^b g(x)dx \approx \int_a^b P(x)dx &= \int_a^b \left(\sum_{k=1}^{N} g(x_k) l_k(x) \right) dx \\ &= \sum_{k=1}^{N} g(x_k) \int_a^b l_k(x)dx = \sum_{k=1}^{N} w_k g(x_k). \end{aligned} \tag{2.11}$$

Anyhow, from the practical point of view, the most interesting aspect is that the definition of the evaluation points, $x_1, x_2, \cdots, x_N$, as well as the weights, $w_1, w_2, \cdots, w_N$, are independent from the function $g(x)$. The only requirement is that it must be possible to evaluate $g(x)$ at the defined $x_1, x_2, \cdots, x_N$. Thus, once $x_1, x_2, \cdots, x_N$ and $w_1, w_2, \cdots, w_N$ have been computed, they can be used to approximate the integral of any function (Haber, 1970).

Accordingly, Haber (1970) states that in order to find N-point quadrature formulas of a high degree of precision, one must solve the following system of equations:

$$\int_0^1 x^r dx = \sum_{k=1}^{N} w_k x_k^r \qquad \text{for } r = 0,1,2,\cdots,d, \tag{2.12}$$

for as high a value of d as possible.

Note that the domain of integration in (2.12) is $[0,1]$ and that the quadratures can be transformed to apply to any interval $[a, b]$. The point x_k for $[0,1]$ may be transformed into $x_k{}'$ for $[a, b]$, as follows (Haber, 1970):

$$x_k' = a + (b - a)x_k. \tag{2.13}$$

The weight w_k must also be transformed into $w_k{}'$ for $[a, b]$, and this can be done with the formula:

$$w_k' = (b - a)w_k. \tag{2.14}$$

With the Newton-Cotes formulas, it is hoped to solve (2.12) for $d \leq N - 1$. However, Gauss showed that by properly choosing the quadrature points, $x_1, x_2, \cdots, x_N$, and weights, $w_1, w_2, \cdots, w_N$, one can solve the system for $d = 2N - 1$, so that there is an N-point formula of the degree of precision of $2N - 1$; and there is no N-point formula of any higher degree of precision (Hildebrand, 1956, as cited by Haber, 1970).

Such an improvement on the efficiency of the quadrature points – in terms of the achieved degree of precision of the formula – can be obtained because the Gaussian quadrature formula relaxes the restriction of equal distance of points of the Newton-Cotes formulas and it determines the optimal distribution of $x_1, x_2, \cdots, x_N$ and $w_1, w_2, \cdots, w_N$; thus, it is possible to supply comparable accuracy with fewer points (Hildebrand, 1987). Another characteristic of the Gaussian integration rules is that these do not interpolate polynomials in the Lagrange form, but they use polynomials belonging to a class of orthogonal polynomials; and it can be shown that the evaluation points are the roots of these polynomials (Stoer and Bulirsch, 2002).

Accuracy of Approximation of the Quadrature Formulas

With respect to the accuracy of approximation of the Newton-Cotes integration formulas with interpolated polynomials in the Lagrange form of degree $N - 1$, Haber (1970) explained that the quadratures have the property that the approximation value is exactly equal to the integral of $g(x)$, if $g(x)$ happens itself to be a polynomial of degree $N - 1$ or lower. The integration formula may or may not be exact for polynomials of higher degree.

With respect to the Gaussian formulas, these have proven to be very accurate in practice (Haber, 1970).

Nevertheless, the accuracy with which the numerical integration formulas – Newton-Cotes and Gaussian quadratures – approximate the integral of $g(x)$

depends not only on the degree of accuracy of the formula but also on the accuracy with which $g(x)$ itself can be approximated by polynomials (Haber, 1970). This depends on the "smoothness" of $g(x)$, which refers to how differentiable (l times differentiable) the integrand is, in the domain of integration $[a, b]$, and whose lth derivative is continuous (Haber, 1970). If we name the approximation value obtained with a quadrature formula: $Q(g(x))$, then we can express its approximation error by (Haber, 1970):

$$\int_a^b g(x)dx - Q(g(x)). \tag{2.15}$$

In addition, since in most of the cases the true value of $\int_a^b g(x)dx$ cannot be computed – because, e.g., a polynomial which describes the function is not available and the function values are only known at some determined points – then, the error of approximation of the quadratures, as in (2.15), cannot be estimated. As a result, in such situations, the true value of $\int_a^b g(x)dx$ is approximated by searching the convergence point of several approximation results computed, commonly, with increasing N. Then, to determine the approximation error of the quadratures, (2.15) is calculated using the approximation of the true value of $\int_a^b g(x)dx$.

The approximation error of Newton-Cotes and of Gaussian formulas can be reduced, in most cases, by increasing N (by increasing the degree of precision of the quadrature). However, if an integrand is not very smooth and the accuracy desired is not achieved easily, "composite" formulas present a good alternative and are very convenient (Hildebrand, 1987). These formulas consist of dividing $[a, b]$ into smaller intervals, applying interpolation formulas to those subintervals and adding the results.

2.1.2 Discrete Approximations of Probability Density Functions

So far, the basic concepts on numerical evaluation of single integrals of the form $\int_a^b g(x)dx$ have been summarized. This section uses those concepts and extends the theoretical analysis to the cases in which the independent variable is a random variable.

In the evaluation of integrals of functions with random variables the objective is to determine its central moments (mean, variance, skewness, kurtosis, etc.). Thus, the numerical integration problem treated in this section is of the form:

$$\int_{-\infty}^{\infty} g(x)f(x)dx, \tag{2.16}$$

with the approximate solution:

$$\int_{-\infty}^{\infty} g(x)f(x)dx, \approx \sum_{k=1}^{N} p_k g(x_k). \tag{2.17}$$

Note that due to the random condition of the variables, now the domain of integration is $[-\infty, \infty]$ and for every realization of the stochastic variable x the PDF, $f(x)$, takes an associated value.

Assuming that the functions of random variables can be approximated by polynomials, Miller and Rice (1983) write equation (2.17) in extended form as follows:

$$\begin{aligned} &\int_{-\infty}^{\infty} (\alpha_0 + \alpha_1 x + \alpha_2 x^2 + \cdots) f(x) d\mathrm{x} \\ &\qquad \approx \sum_{k=1}^{N} p_k (\alpha_0 + \alpha_1 x + \alpha_2 x^2 + \cdots). \end{aligned} \tag{2.18}$$

They continue to analyze the mathematical problem and explain that (2.18) can be rewritten in terms of the moments of the original and approximate distribution as follows:

$$\begin{aligned} &\alpha_0 + \alpha_1 \mathrm{E}[x] + \alpha_2 \mathrm{E}[x^2] + \cdots \\ &\qquad \approx \alpha_0 \sum_{k=1}^{N} p_k + \alpha_1 \sum_{k=1}^{N} (p_k\, x_k) + \alpha_2 \sum_{k=1}^{N} (p_k x_k^2) + \cdots; \end{aligned} \tag{2.19}$$

and that (2.19) can be satisfied for any set of coefficients $(\alpha_0, \alpha_1, \alpha_2, \cdots)$ if we make the moments of the original distribution of the random variable equal to the moments of the approximation.

Therefore, the criterion for the approximation becomes (Miller and Rice, 1983):

$$\text{(2.20)} \qquad \mathrm{E}[x^r] = \int_{-\infty}^{\infty} x^r f(x)\, dx = \sum_{k=1}^{N} p_k x_k^r \qquad \text{for } r = 0,1,2,\cdots,d.$$

Now, if we extend the system of equations from (2.20) and consider the moments of the distribution as central moments, we obtain the following (own development based on Artavia et al., 2009):

$$\text{(2.21)} \quad \begin{aligned}
&\text{Measure of } [-\infty,\infty]: && \mathrm{E}[x^0] = \int_{-\infty}^{\infty} \{x\}dx = 1 = p_1 + p_2 + \cdots + p_N, \\
&\text{Mean of } x: && \mathrm{E}[x^1] = \int_{-\infty}^{\infty} x\{x\}dx = p_1x_1 + p_2x_2 + \cdots + p_Nx_N, \\
&\text{Variance of } x: && \mathrm{E}[(x-\mathrm{E}[x])^2] = \int_{-\infty}^{\infty} (x-\mathrm{E}[x])^2 f(x)dx \\
& && \quad = p_1(x-\mathrm{E}[x])^2 + \cdots + p_N(x-\mathrm{E}[x])^2, \\
& && \quad \vdots \\
&\text{Higher central moments:} && (x-\mathrm{E}[x])^d = \int_{-\infty}^{\infty} (x-\mathrm{E}[x])^d f(x)dx \\
& && \quad = p_1(x-\mathrm{E}[x])^d + \cdots \\
& && \quad + p_N(x-\mathrm{E}[x])^d.
\end{aligned}$$

In this way, it can be observed that the criterion for the approximation is to generate sets of points and probabilities which match the central moments of the PDF of the random variable to a certain desired degree of precision. Accordingly, Miller and Rice (1983) named these quadrature formulas: "discrete approximations of probability distributions". Moreover, in order to comprehend the correspondence between the numerical integration formulae from the last section and the discrete approximation of probability distributions, we associate the PDF, $f(x)$, with the weighting function, $w(x)$, and the probabilities, p_k, with the weights, w_k (Miller and Rice, 1983).

For example, notice that in the case of random variables with uniform distribution in the interval [0,1], the PDF is:

$$f(x) = \frac{1}{1-0} = 1 \qquad \text{for } 0 \leq x \leq 1, \tag{2.22}$$

and then, by substituting (2.22) in (2.20), we obtain:

$$\mathrm{E}[x^r] = \int_0^1 x^r dx = \sum_{k=1}^{N} p_k x_k^r \qquad \text{for } r = 0,1,2,\cdots,d, \tag{2.23}$$

which is exactly equal to the system of equations (2.12) exposed by Haber (1970). This demonstrates that one can interpret the numerical integration formulas from Section 2.1.1 as discrete approximations of the moments of uniform distributions of x for the interval [0,1]; which then through appropriate transformation approximates any desired interval $[a, b]$. Additionally, when $r = 0$ the system of equations (2.12) and (2.21) give the measure of the interval of integration and the measure of the PDF of x respectively; thus, we may also make a correspondence between the domain of integration of numerical integration formulas and the PDF of *x* of discrete approximations of probability distributions. As a result, we may say that the criterion to find numerical integration formulas is to generate sets of points and weights which match the moments of the domain of integration; and that the criterion to find discrete approximations of probability distributions is to generate sets of points and probabilities which match the moments of the PDF of x.

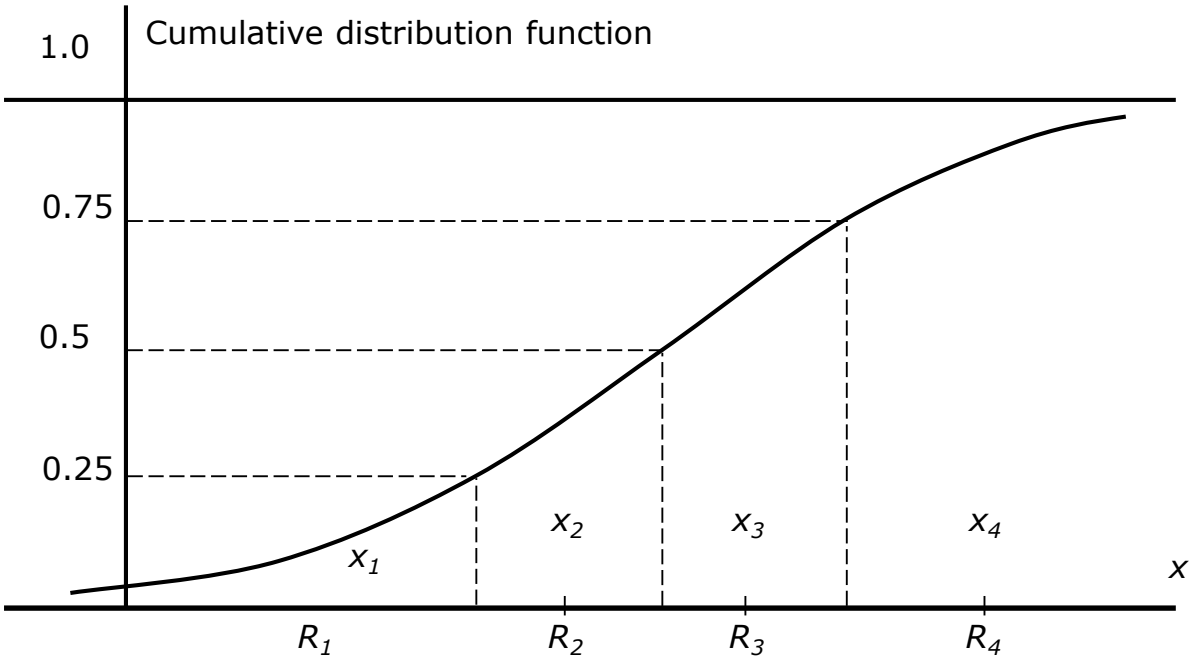
Commonly Used Approximation Techniques

Miller and Rice (1983) explain how discrete approximations of probability distributions obtained by commonly used approximation techniques not always match the moments of the original probability distribution and propose the Gaussian quadrature method of numerical integration as a more accurate alternative.

In their paper, typical techniques are defined as procedures which divide the range of possible values or range of cumulative probabilities into a set of collectively exhaustive and mutually exclusive intervals. Also, in these procedures, each interval is approximated by a value equal to its mean or median and a probability equal to the chance that the true value is in the interval. The approach is shown in Figure 2.1, where the range of cumulative probabilities is

subdivided into four equal segments and the corresponding regions, R_1, R_2, R_3 and R_4, on the x-axis are represented by their means.

Figure 2.1: Approximating a probability distribution by means of equally likely intervals



Source: Miller and Rice (1983, p. 353)

Miller and Rice (1983) also explain that when applying these commonly used approximation techniques, the variance and other higher central moments will usually be underestimated. To show this, they approximate several well-known distributions using the procedure illustrated in Figure 2.1 and compute the percentage error (see Table 2.1).

In Table 2.1, it can be observed how the variances of the different distributions are, indeed, underestimated. Also, it can be seen that, as with other quadrature methods, when N is larger the approximation is more accurate.

Finally, it is argued that these errors could actually be acceptable given the lack of precision of the subjective probability distributions used as inputs; nonetheless, these errors would not be random, and the approximation will systematically underestimate the degree of variability inherent in the original distributions (Miller and Rice, 1983)

Table 2.1: The percentage error in the central moments of discrete approximations of several distributions using the means of equally likely intervals

		Approximation error in the following central moments in %:			
		Mean	Variance	Skew	Kurtosis
Distribution	N	$\mathrm{E}[x]$	$\mathrm{E}[(x-\mathrm{E}[x])^2]$	$\mathrm{E}[(x-\mathrm{E}[x])^3]$	$\mathrm{E}[(x-\mathrm{E}[x])^4]$
Uniform	2	0.0	-25.0	0.0	-68.7
$f(x)=1:0\le x\le 1$	3	0.0	-11.1	0.0	-34.2
	4	0.0	-6.2	0.0	-19.9
Beta	2	0.0	-31.5	-100.0	-80.1
$f(x)=12x(1-x)^2 :$ $0\le x\le 1$	3	0.0	-16.0	-76.7	-55.5
	4	0.0	-9.9	-62.0	-40.3
Normal	2	0.0	-36.3	0.0	-86.5
$f(x)=\left(\frac{1}{\sqrt{2\pi}}\right)e^{-\left(\frac{1}{2}\right)x^2}$	3	0.0	-20.7	0.0	-68.5
	4	0.0	-13.9	0.0	-56.3
Exponential	2	0.0	-52.0	-100.0	-97.4
$f(x)=e^{-x}:0\le x$	3	0.0	-35.1	-87.2	-93.0
	4	0.0	-26.5	-76.5	-88.0
Binomial	2	0.0	-36.1	-291.9	-82.1
$f(x)=$ $\binom{10}{x}(0.4)^x(0.6)^{(10-x)}$	3	0.0	-19.4	-73.4	-69.1
$:x=1,2,\cdots,10$	4	0.0	-9.7	-80.7	-34.2

(N= Number of probability-value pairs in the discrete approximation)

Source: Miller and Rice (1983, p.354)

2.2 Numerical Evaluation of Multiple Integrals

2.2.1 Basic Concepts Continued

After dealing with the concepts of numerical evaluation of integrals of univariate functions, $\int_a^b g(x)\,dx$ and $\int_{-\infty}^{\infty} g(x)f(x)dx$, in the last two sections, the theory concerned with the numerical evaluation of definite integrals of multivariate functions is studied in this section. These mathematical problems are expressed as follows:

$$\int \cdots \int_R g(x_1, x_2, \cdots, x_n)dx_1 dx_2 \cdots dx_n, \tag{2.24}$$

or shortly written as:

$$\int_R g(\mathbf{x})d\mathbf{x}. \tag{2.25}$$

Note that many of the developments set out here are analogous to those of Section 2.1.1 but extended in order to account for the additional complexity due to multidimensionality.

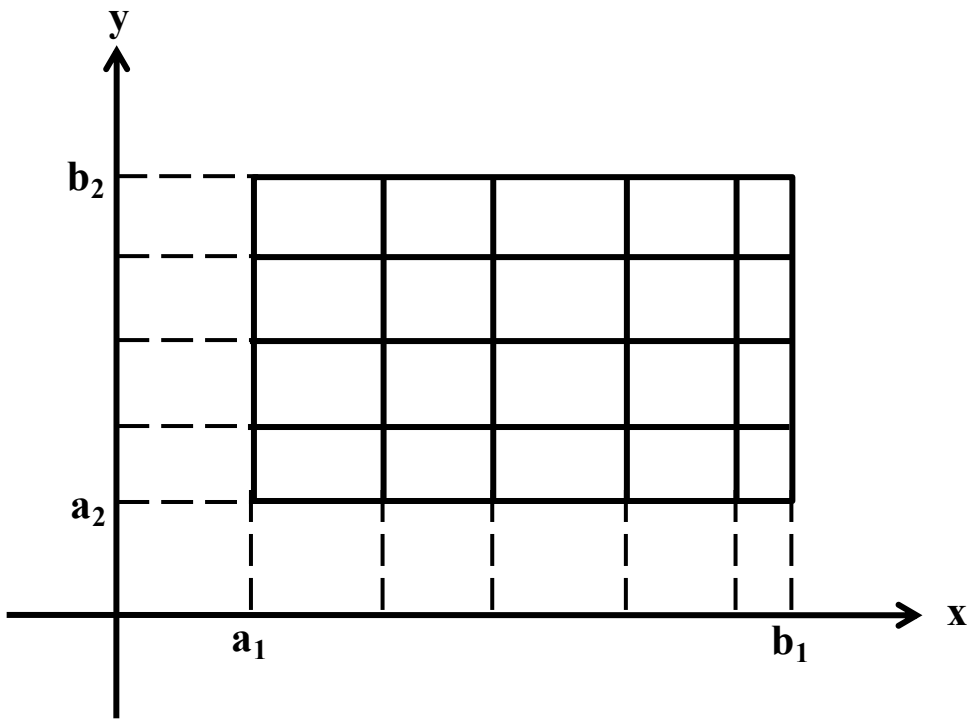
First of all, the numerical evaluation of multiple integrals presents the characteristic that there are infinitely many regions that are not equivalent under affine transformations (Haber, 1970). Due to this characteristic in two or more dimensions, it appears that one needs a new numerical quadrature theory for each affine class of regions (Haber, 1970). In contrast, remember that for the 1-dimensional case one can apply the transformations from (2.13) and (2.14), making the quadrature valid for any interval of integration. Therefore, in problems concerned with multidimensionality and in order to keep track of the theory, the approach of researchers has been to develop formulas for the simplest regions (e.g., n-cube, n-sphere, n-simplex); as it is expected that these regions will most commonly occur in practice (Haber, 1970).

In the same manner as for the 1-dimensional case, the basic concept of the numerical integration methods for multiple integrals is to solve the integration problem using discretization methods which approximate the integral by finite sums corresponding to some partition of the domain of integration. However, as already seen from (2.24) and (2.25), the domain of integration for multiple integrals is represented by regions, R, and not by intervals, $[a, b]$.

In order to imagine how this works, a simple case is presented:

Let $g(\mathbf{x})$ be a function defined on a rectangle, H^2, and let $P = \{H_1^2, H_2^2, \cdots, H_N^2\}$ be a partition of H^2.

Figure 2.2: Illustration of a partitioned rectangle H^2



Source: Trench (2011, p.438)

Now, let the content of any H_k^2 for $k = \{1,2,\cdots,N\}$ be denoted by $V(H_k^2)$, which in this case gives the area of that unit of the partition. For 3-dimensional problems, $V(H_k^3)$ gives the volume of the partition unit k; for higher dimensions, $V(H_k^n)$ gives the n-volume. Finally, let $\mathbf{x}_k$ be an arbitrary point in H_k^2.

Then (Trench 2011):

$$\int_{H^2} g(\mathbf{x})d\mathbf{x} \approx \sum_{k=1}^{N} \mathrm{V}(H_k^2)\, g(\mathbf{x}_k) \tag{2.26}$$

is an approximation of the multiple integral of $g(\mathbf{x})$ over H^2 obtained by a Riemann sum.

This kind of quadrature has the problem that, with few evaluations, approximations will often be of low accuracy. Nonetheless, the number of evaluations, N, can be gradually increased until convergence of the results is achieved. However, the researcher must be aware that for high dimensional problems and for complex function forms, e.g., described by polynomials of high degree, the number of evaluations required for convergence may be very high.

The most obvious approach to solve multiple integral problems is to treat the integral as an n-fold iterated integral – to integrate over manifolds as the n-cube,

n-simplex, etc. – and to apply the basic 1-dimensional integration rules to each variable separately (Haber, 1970). For example, a 2-dimensional integral over the n-cube G^2 is approximated by (own development based on Haber, 1970):

$$\int_{G^2} g(x,y)dxdy = \int_0^1 \left(\int_0^1 g(x,y)dx \right) dy \approx \int_0^1 \big(Q_x(g;x)\big)\, dy \approx Q_y(Q_x(g;x);y), \tag{2.27}$$

where the expression $Q_x(g;x)$ refers to the quadrature of $g(x,y)$ with respect to x and $Q_y(Q_x(g;x);y)$ indicates the quadrature of $g(x,y)$ with respect to x and y. With $Q_y(Q_x(g;x);y)$, the function, $g(x,y)$, is evaluated several times at the quadrature points in the y-axis, using every time the different quadrature points in the x-axis.

For the generalization of (2.27) to higher dimensions, its expressions are simplified to: $Q_i(g)$ and $Q_j\big(Q_i(g)\big)$. Then, for the n-dimensional integral over G^n with $Q_1, Q_2, \cdots, Q_n$ being the quadrature formulas applied to $g(\mathbf{x})$ with respect to each variable, the multiple quadrature may be expressed as follows (adapted from Haber, 1970):

$$\begin{aligned} \int_{G^n} g(\mathbf{x})d\mathbf{x} &= \int_0^1 \cdots \int_0^1 g(x_1, x_2, \cdots, x_n)dx_1dx_2\cdots dx_n \\ &\approx \int_0^1 \cdots \int_0^1 \big(Q_1(g)\big)dx_2dx_3\cdots dx_n \\ &\approx \int_0^1 \cdots \int_0^1 \Big(Q_2\big(Q_1(g)\big)\Big)\, dx_3\cdots dx_n \\ &\approx Q_n\Big(Q_{n-1}\Big(\cdots\big(Q_1(g)\big)\Big)\Big). \end{aligned} \tag{2.28}$$

Now, if:

$$Q_i(g) = \sum_{k_i=1}^{N_i} w_{k_i} g\big(x_{k_i}\big) \qquad \text{for } i = 1,2,\cdots,n, \tag{2.29}$$

then, the approximation of the multiple integral can be rewritten as follows (adapted from Haber, 1970):

$$\int_{G^n} g(\mathbf{x})d\mathbf{x} \approx \sum_{k_1=1}^{N_1}\sum_{k_2=1}^{N_2}\cdots\sum_{k_n=1}^{N_n} w_{k_1}w_{k_2}\cdots w_{k_n} g\left(x_{k_1}, x_{k_2}, \cdots, x_{k_n}\right). \tag{2.30}$$

Equation (2.30) indicates that for the approximation, the multivariate function is evaluated at all the possible combinations of the quadrature points generated for each variable. In other words, the function is evaluated at the vector points obtained from the Cartesian product of the set of points of every quadrature. Note that the weight for every vector point is obtained from the multiplication of the weights for that point for each variable. Accordingly, these multivariate quadratures are often called "Cartesian product" formulas.

With respect to the degree of precision of the Cartesian product quadratures, if, for each i, d_i is the degree of precision of Q_i, then the degree of precision of the multivariate quadrature is $min\{d_1, d_2, \cdots, d_n\}$ (Haber, 1970).

As it may have already been seen, the Cartesian product technique requires many more evaluations of the integrand than the basic 1-dimensional quadrature rules in order to obtain the same degree of precision. For example, if the same 1-dimensional quadrature, Q, is used for each i and $\widetilde{N}$ is the number of evaluations required by Q, then the total number of evaluations required, N, is (adapted from Haber, 1970):

$$N = \widetilde{N}^n \tag{2.31}$$

Therefore, if it has been decided to use a 10-point 1-dimensional quadrature formula, then the 2-, 3- and 4-dimensional case, using the Cartesian product formulas, requires 100, 1000 and 10000 points, respectively.

If the integrand for some reason is costly to evaluate, e.g., in terms of time per evaluation and the dimension of the problem is relatively high, then more efficient quadrature formulas will be required. In this sense, analogous to (2.12), Haber (1970) stated that the approach to finding more efficient quadrature formulas of a given degree, d, of precision is to solve (2.32) using the fewest possible points.

$$\int_R \mathbf{x}^{\mathbf{r}} d\mathbf{x} = \sum_{k=1}^{N} w_k \mathbf{x}_k^{\mathbf{r}} \qquad \text{for } |\mathbf{r}| = 0,1,2,\cdots,\mathrm{d}. \tag{2.32}$$

In particular, for $|\mathbf{r}| = 0,1,2$, this means:

(2.33) $$|\mathbf{r}| = 0: \quad \int_R d\mathbf{x} = \sum_{k=1}^{N} w_k,$$

$$|\mathbf{r}| = 1: \quad \int_R x_i d\mathbf{x} = \sum_{k=1}^{N} w_k \, x_{k,i} \qquad \text{for all } i = 1,2,\cdots,n,$$

$$|\mathbf{r}| = 2: \quad \int_R x_{i_1} x_{i_2} d\mathbf{x} = \sum_{k=1}^{N} w_k \, x_{k,i_1} x_{k,i_2}$$

$$\text{for all } i_1, i_2 = 1,2,\cdots,n.$$

For their aim of using the fewest possible points, these formulas are called throughout this study as 'efficient quadratures'.

Anyhow, notice that actually every quadrature formula – efficient or not – must satisfy the system of equations presented in (2.32 and 2.33). This is a requirement because, with the generated points and weights of any formula, the objective is to approximate the moments of the region of integration to a certain degree, d, of precision; which is exactly what (2.32) does.

Therefore, the weights and points generated with Cartesian product formulas will also satisfy (2.32) to a certain degree, d. The difference to 'efficient quadratures' is that these will solve the system, for the same d, with a much smaller N. The simplest example of an efficient quadrature formula is the fact that for any integration domain there is a formula of degree 1 of 1 point. In that case, the point would be the centroid of the region, and the weight would be the measure $|R|$ of it (Haber, 1970).

A basic approach used in the search for efficient formulas is based on the symmetry properties of certain integration regions (Haber, 1970). This occurs because for fully symmetrical regions as the n-cube and the n-sphere it is easier to obtain a certain degree, d, of precision with few points. Normally, these formulas place the center of the region at the origin of the coordinate system and generate points which are distributed symmetrically around that origin. Nonetheless, the system of equations of (2.32) may be solved in different manners. For example, it may happen that some formulas generate evaluation points which lie outside the region of integration. According to Haber (1970), in the earliest paper on efficient multiple quadratures from Maxwell (1877) one formula used such points. Anyhow, these formulas present the problem that in some cases the functions under consideration may not be defined outside the domain of integration (Haber, 1970).

Many efficient multidimensional quadrature formulas of different degree of precision and for different integration regions can be found in the literature. Haber (1970) summarized the main findings up to 1970. These were composed of the following:

1. Formulas of degree 2 with $N = n + 1$ for any n-dimensional region from Thacher (1957) and Stroud (1960). Their number of points is minimal but for most of the formulas these lie outside of R.
2. A formula of degree 3 with $N = 2n$ for C^n from Tyler (1953). This formula uses points on the coordinate axes and has the points outside R for $n > 3$ (Stroud, 1957).
3. The formula of degree 3 with $N = 2n$ for C^n from Stroud (1957). In this case, the points do not lie on the coordinate axes and all points are interior to R for all n (Stroud, 1957). Mysovskih (1966) (as cited by Haber, 1970) proved that the number of points used in this formula is minimal.
4. A formula of degree 3 with $N = n + 2$ for integrating over the n-dimensional simplex from Hammer and Stroud (1956).
5. A formula of degree 4 with $N = 6$ for G^2 and B^2 from Mysovskih (1968) (as cited by Haber, 1970) with the minimum possible number of points.

Schürer (2008), in a manual for HIntLib – a software containing a library of methods for high-dimensional numerical integration –, states that the most comprehensive collection of rules has been compiled by Stroud (1971) and that this work was continued by Cools and Rabinowitz (1993). Out of these two compilation works, Schürer (2008) selects and presents the rules for the integration over the hypercube that contains all the information for actually implementing them. He summarizes these rules in Chapter 7 of his manual.

As for the 1-dimensional case, the accuracy with which Cartesian product formulas and efficient quadrature formulas approximate multiple integrals depends not only on the degree of precision, d, of the quadrature formula but also on the smoothness of $g(\mathbf{x})$.

2.2.2 Monte Carlo Approach

There is one numerical evaluation method for which the smoothness of $g(\mathbf{x})$ is not essential: the Monte Carlo approach. Its basic idea is to contemplate the evaluation of the multiple integral as a probabilistic problem and to investigate it by statistical experiments (Metropolis and Ulam, 1949). In practice, it generates samples of the integration region – which may be considered as the homologous to the input PDF in probabilistic problems (see Section 2.1.2) – of equally

weighted random points and approximates the integral by the following formula (Haber, 1970):

$$\int_R g(\mathbf{x}) \approx \sum_{k=1}^{N} w_k \, g(\mathbf{x}_k), \tag{2.34}$$

where (Haber, 1970):

$$w_k = \frac{|R|}{N}. \tag{2.35}$$

Notice that the Monte Carlo approach solves the integration problem as all other numerical integration formulas – using discretization methods which approximate the integral by finite sums corresponding to some partition of the domain of integration. Nonetheless, the difference of Monte Carlo quadratures with respect to, e.g., Newton-Cotes and Gaussian quadratures, is that it is not based on interpolation approaches and that the evaluation points $\mathbf{x}_k$ are determined randomly.

In terms of accuracy of approximation (see Equation 2.15 for the definition of approximation error), in order to obtain a desired accuracy level, the random characteristic of the Monte Carlo approach results in much higher computational requirements when compared to Newton-Cotes or Gaussian quadratures. For example, Haber (1970) asseverates that even though the Monte Carlo method has the advantage that it can be used for functions with singularities (non-continuous lth derivatives), it is almost never used in one dimensional problems, since the desired accuracy level is obtained faster with numerical integration methods based on interpolation. In these cases, if the functions present some singularities – which generally occur at isolated points –, these may be gotten around by various devices (Haber, 1970). For example, composite formulas may be convenient for the numerical evaluation of this kind of 1-dimensional functions.

Conversely to the single variable case, for functions of multiple dimensions, singularities occur along surfaces or manifolds of complicated shape and can rarely be removed (Haber, 1970). Therefore, the Monte Carlo method is a very attractive tool for the handling of multidimensional integration problems of functions $g(\mathbf{x})$ with singularities. Also, the method presents the advantage that it can be used for the approximation of integrals over any n-dimensional region – remember that with other numerical integration methods it appears that one needs a new quadrature theory for each affine class of regions (see Section 2.2.1).

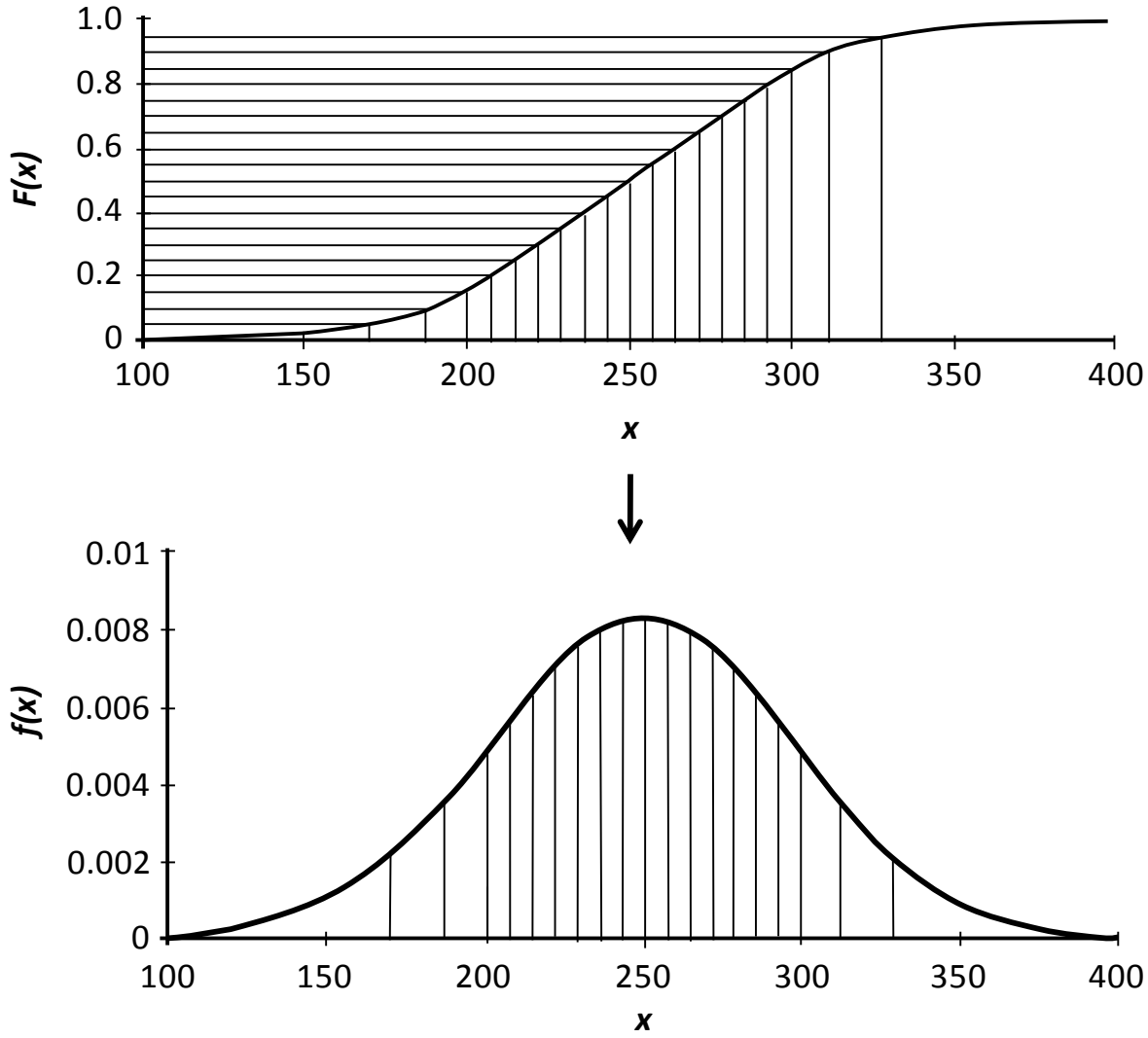
Currently, several sampling procedures may be classified as Monte Carlo methods. Here, the two most common sampling methods (Vose, 2000): the Monte Carlo Sampling (MCS) and the Latin Hypercube Sampling (LHS) are presented.

Briefly explained, the MCS provides equal opportunity of an **x**-value being generated in any percentile range of the input distribution – in the case of numerical integration formulas, the region of integration may be considered to be uniformly distributed (see Section 2.1.2). Conversely, the LHS uses a technique which divides first the input distribution in N intervals of equal probability and then generates one random x_k for every interval k (Vose, 2000). In the multivariate case, the LHS would divide (partition) the n-dimensional region into n-dimensional sub regions of equal probability and would generate one random $\mathbf{x}_k$ for every sub region.

Note that the LHS is similar to the commonly used approximation techniques described by Miller and Rice (1983) (see Section 2.1.2). In both cases, the probability distribution is divided into intervals of equal probability; nevertheless, on the contrary to commonly used techniques, the LHS approximates each interval by a random value of it and not by its mean or median.

Figure 2.3 exemplifies the effect of the stratification produced for 20 iterations ($k = 20$) in the LHS. It can be observed how the bands get progressively wider when approaching the tails of the distribution where the probability density drops (Vose, 2000).

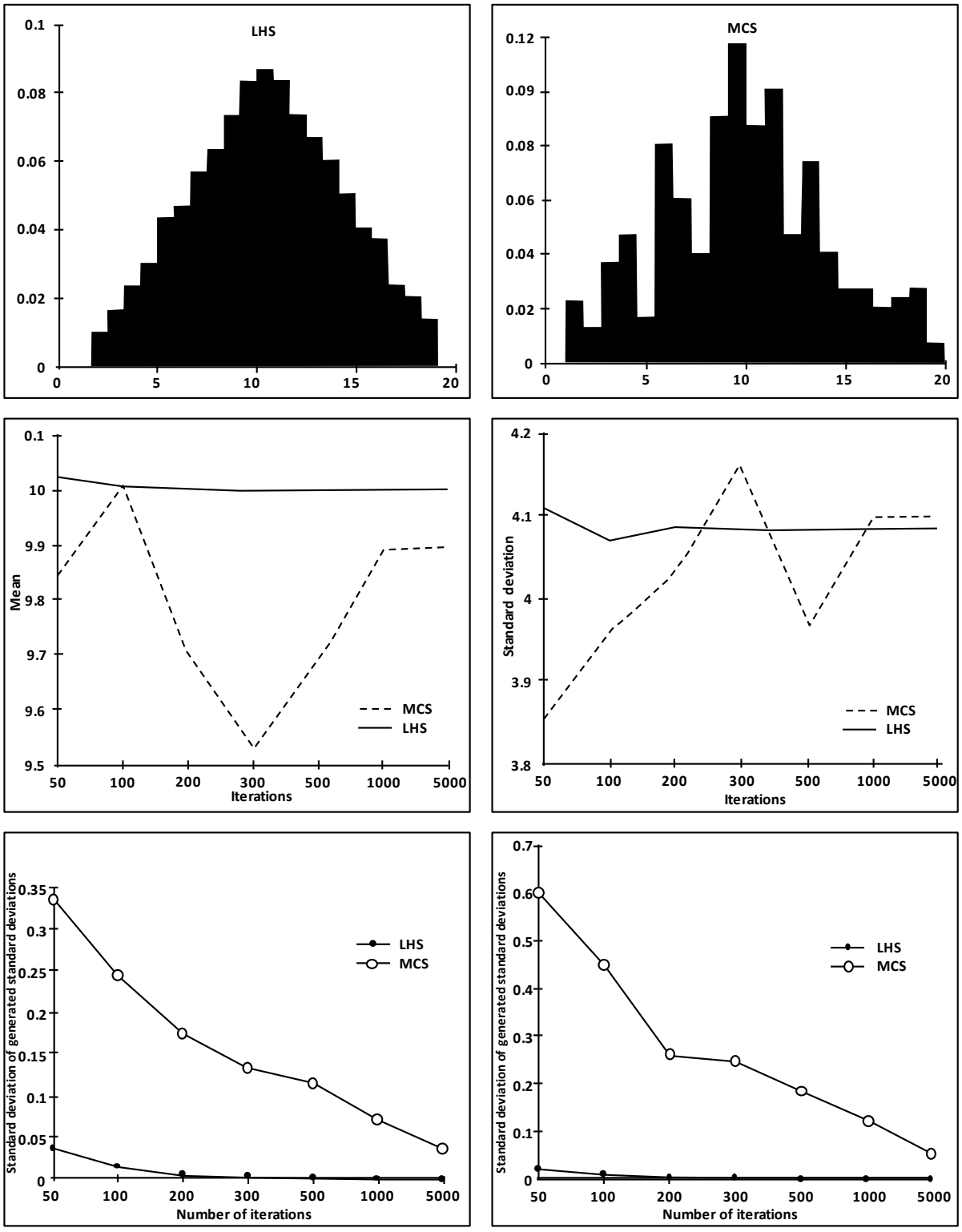
Figure 2.3: Example of the effect of stratification in LHS



Source: Vose (2000, p. 60)

Further below, Figure 2.4 demonstrates the improvement achieved by LHS when compared to MCS. The upper panel of the figure shows the histograms obtained after one simulation of 300 iterations ($N = 300$) of a triangular distribution $Triang(0,10,20)$. The middle panel gives the mean and the standard deviation obtained with both methods with progressively increasing N. In it, it can be clearly seen how convergence is obtained much faster with LHS than with MCS. The lower panel presents, on the left, the standard deviations of means and, on the right, the standard deviation of standard deviations, both obtained after computing 100 simulations for every sample size ($N = 50, N = 100, \cdots, N = 5000$). In this last panel, it can be observed that with MCS the mean and the standard deviation of different samples of same size may vary significantly, while different samples of same size generated with LHS are much more similar to each other.

Figure 2.4: Comparison of the performance of Latin Hypercube Sampling (LHS) and Monte Carlo Sampling (MCS)



Source: Vose (2000, p. 60)

2.2.3 Discrete Approximations of Multivariate Probability Density Functions

Sections 2.2.1 and 2.2.2 present the basic concepts of numerical integration methods for the evaluation of multiple integrals – integrals of multivariate functions – over determined regions of integration, R. This section extends the analysis to the case of functions dependent on multivariate random variables. Many expositions are the generalization, to the multivariate case, of the statements from Section 2.1.2.

As with the integral of functions of single random variables, with multivariate random problems, we are interested in determining the central moments of the multivariate function $g(\mathbf{x})$, e.g.:

$$\mathrm{E}[\, g(\mathbf{x})] = \int_{\mathbb{R}^n} g(\mathbf{x})\, f(\mathbf{x})\, d\mathbf{x} \tag{2.36}$$

and:

$$\mathrm{E}[(\, g(\mathbf{x}) - \mathrm{E}[g(\mathbf{x})])^2] = \int_{\mathbb{R}^n} (g(\mathbf{x}) - \mathrm{E}[g(\mathbf{x})])^2 f(\mathbf{x})\, d\mathbf{x}, \tag{2.37}$$

which are the mean and the variance of $g(\mathbf{x})$.

The solution to these problems is given by:

$$\int_{\mathbb{R}^n} g(\mathbf{x})\, f(\mathbf{x})\, d\mathbf{x} \approx \sum_{k=1}^{N} p_k g(\mathbf{x}_k). \tag{2.38}$$

Note that due to the random condition of the multivariate integration problem, now the domain of integration is the n-dimensional Euclidian space, $\mathbb{R}^n$, and that every evaluation of $g(\mathbf{x})$ has an associated probability, which is given by $f(\mathbf{x})$. Consequently, in the approximation, every evaluation point, $\mathbf{x}_k = \left(x_{k,1}, x_{k,2}, \cdots, x_{k,n}\right)$ for $k = 1,2, \cdots, N$, also has a corresponding probability, p_k.

As in Section 2.1.2, assuming that the function $g(\mathbf{x})$ can be approximated by polynomials, the criterion for the approximation becomes (own development based on Miller and Rice, 1983 and Artavia et al., 2009):

$$\mathbf{x}^{\mathrm{r}} = \int_{\mathbb{R}^n} \mathbf{x}^{\mathrm{r}} f(\mathbf{x}) d\mathbf{x} = \sum_{k=1}^{N} p_k \mathbf{x}_k^r \qquad \text{for } |\mathbf{r}| = 0,1,2, \cdots, d, \tag{2.39}$$

where $\mathbf{x}^{\mathrm{r}}$ is a monomial in n variables of degree $|\mathbf{r}|$.

Now, if we expand (2.39) for $|\mathbf{r}| = 0,1,2$, we get the following system of equations (own development based on Artavia et al., 2009):

(2.40)
$$|\mathbf{r}| = 0: \quad \int_{\mathbb{R}^n} f(\mathbf{x})d\mathbf{x} = \sum_{k=1}^{N} p_k = 1,$$

$$|\mathbf{r}| = 1: \quad \int_{\mathbb{R}^n} x_i f(\mathbf{x})d\mathbf{x} = \sum_{k=1}^{N} p_k\, x_{k,i} = \mu_i \qquad \text{for all } i = 1,2,\cdots,n$$

$$\Leftrightarrow \quad \int_{\mathbb{R}^n} \mathbf{x} f(\mathbf{x})d\mathbf{x} = \sum_{k=1}^{N} p_k\, \mathbf{x}_k = \boldsymbol{\mu},$$

$$|\mathbf{r}| = 2: \quad \int_{\mathbb{R}^n} \mathrm{E}\big[(x_i - \mathrm{E}[x_i])\big(x_j - \mathrm{E}\big[x_j\big]\big)\big] f(\mathbf{x})d\mathbf{x}$$
$$= \sum_{k=1}^{N} p_k \big(x_{k,i} - \mathrm{E}[x_i]\big)\big(x_{k,j} - \mathrm{E}\big[x_j\big]\big)$$
$$= \sigma_{i,j}$$

for all i, j where $i = j = 1,2,\cdots,n$

$$= \begin{bmatrix} \sigma_{1,1} & \cdots & \sigma_{1,n} \\ \vdots & \ddots & \vdots \\ \sigma_{n,1} & \cdots & \sigma_{n,n} \end{bmatrix} = \boldsymbol{\Sigma}\,.$$

In (2.40), it can be observed that the purpose of the system of equations is to produce discrete approximations of the multivariate PDF of $\mathbf{x}$ – to find sets of probabilities and vector points, $(p_k, \mathbf{x}_k)$ – whose moments equal the moments of the original PDF to a desired degree of precision, $|\mathbf{r}| = d$.

For $d = 3$ the following central moments are approximated:

with $|\mathbf{r}| = 0$: the measure of the multivariate PDF,

with $|\mathbf{r}| = 1$: the vector of means, $\boldsymbol{\mu}$,

with $|\mathbf{r}| = 2$: the covariance matrix, $\boldsymbol{\Sigma}$,

with $|\mathbf{r}| = 3$: the skewness and coskewness.

Thus, in (2.40), it can also be observed that not only the moments of the PDF of every stochastic variable alone – of its marginal PDF[5] – must be matched with the discrete approximations, but also the cross moments – the moments involving multiple variables, e.g. covariance and coskewness.

The system of equations in (2.40) may be solved using the efficient formulas presented in Section 2.2.1, provided that the symmetry conditions of the region R for which these were created also apply for the domain of integration in the random case. The system may also be solved with MCS or LHS, which are good alternative method for the approximation of integrals of multivariate functions with singularities.

In stochastic modeling of agricultural market models, random variables are often assumed to be normally distributed. In the single variable case, the normal distribution is expressed by $x \sim N(\mu, \sigma^2)$ with mean μ and variance σ^2. Its PDF is given by (Hinderer, 1972):

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}. \tag{2.41}$$

In the multivariate case, the normal distribution can be written as $\mathbf{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ and its PDF is given by (Hinderer, 1972):

$$f(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n \det(\boldsymbol{\Sigma})}} e^{\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right]}. \tag{2.42}$$

As a result, with normally distributed multivariate random variables, the integration problem presented in equations (2.36) and (2.37) can be approximated by finding solutions to the following system of equations:

$$\mathrm{E}[\mathbf{x}^{\mathrm{r}}] = \int_{\mathbb{R}^n} \mathbf{x}^{\mathbf{r}} \frac{1}{\sqrt{(2\pi)^n \det(\boldsymbol{\Sigma})}} e^{\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right]} d\mathbf{x} = \sum_{k=1}^{N} p_k \mathbf{x}_k^r \tag{2.43}$$

$$\text{for } |\mathbf{r}| = 0,1,2,\cdots,d.$$

[5] This concept of marginal PDF is also used by Gujarati (2003).

3 Stroud's Octahedron for the Approximation of Multivariate Normal Distributions

In the last chapter, it was explained how the moments of the probability distribution of model results can be approximated using numerical integration methods. Furthermore, it has been shown that, especially in multidimensional problems, efficient quadrature formulas of degree $|\mathbf{r}|$ can reduce considerably the required number of evaluations of an integrand – when compared to Cartesian product formulas or to Monte Carlo approaches of numerical integration with the same degree $|\mathbf{r}|$ of precision. The formulas from Arndt (1996) and Artavia et al. (2009) give 'efficient quadratures' for the multivariate standard normal distributions, which are based on Stroud's theorem from 1957 and are thus, quadrature formulas of degree 3 of precision with minimal points. These points can be used as the basis for the generation of quadratures for multivariate normal distributions of the form $\mathbf{x}\sim N(\mathbf{0},\boldsymbol{\Sigma})$ or $\mathbf{x}\sim N(\boldsymbol{\mu},\boldsymbol{\Sigma})$ which are often used in stochastic modeling exercises.

In this chapter, the quadrature formulas from Arndt and Artavia et al. are analyzed in detail with the purpose of understanding how these approximate the regions of integration, as well as determining the properties of each one.

Furthermore, the topic of how to induce a desired correlation to the quadratures is treated theoretically in Section 3.2 and the "how to do it" is treated practically in Section 3.3. The last, gives a detailed example of the computation procedure required for the generation of the quadratures with the desired covariance matrix (including the LHS method). The example highlights some important differences between the approaches shown.

3.1 Quadratures for the Multivariate Standard Normal PDF

3.1.1 Stroud's Theorem and Formula for Symmetric Regions

Stroud's theorem from 1957 for the approximation of multivariate integrals stated the following:

'*A necessary and sufficient condition that* 2n *points* $\mathbf{x}_1, \cdots, \mathbf{x}_n, -\mathbf{x}_1, \cdots, -\mathbf{x}_n$ *form an equally weighted numerical integration formula of degree 3 for a symmetrical region is that these points form the vertices of a* Q^n *whose centroid coincides with the centroid of the region and lie on an* n*-sphere of radius* $r = \sqrt{nI_2/I_0}$' (Stroud, 1957, p. 259).

Where I_0 is the n-volume of the symmetrical region and I_2 is the integral of the square of any of the variables of the region over the entire region.

Furthermore, as stated in Section 2.2.1, it is desirable that the integration formula generates quadrature points that are interior to the region of integration.

Stroud (1957), based on his theorem and with the purpose of producing quadrature points that are interior to the integration region, proposed the following formula for the C^n with vertices $(\pm 1, \pm 1, \cdots, \pm 1)$ (see Box 3.1).

Box 3.1: Stroud's degree 3 quadrature formula for the *n*-cube with vertices $(\pm 1, \pm 1, ..., \pm 1)$ for which the points are interior for all *n*

Let $\boldsymbol{\gamma}_k$ denote the point $(\gamma_1, \gamma_2, \cdots, \gamma_n)$, where:

$$\gamma_{2r-1} = \sqrt{\frac{2}{3}} \cos \frac{(2r-1)k\pi}{n} \qquad \gamma_{2r} = \sqrt{\frac{2}{3}} \sin \frac{(2r-1)k\pi}{n}$$

$$\text{for } r = 1,2, \cdots, \left[\frac{1}{2}n\right]$$

($\left[\frac{1}{2}n\right]$ is the greatest integer not exceeding $\frac{1}{2}n$), and if n is odd:

$$\gamma_n = (-1)^k / \sqrt{3}.$$

Then, the quadrature points $\boldsymbol{\gamma}_1, \cdots, \boldsymbol{\gamma}_N$ for $N = 2n$, satisfy the conditions of the theorem and all are interior to C^n.

Source: Adapted from Stroud (1957, p. 260)

Note that Stroud was confronted with the problem that for the C^n with vertices $(\pm a, \pm a, \cdots, \pm a)$, and for $n > 3$, the radius of the n-sphere on which the vertices of the Q^n must lie on is greater than a. Put another way, for the C^n, for $n > 3$, we get $r > a$ (see Box 3.2). As a result, if the quadrature points lie on the coordinate axes, these will be outside C^n for $n > 3$.

Box 3.2: The radius for the *n*-cube with vertices (±a, ±a,..., ±a)

Region of integration:

$$R = C^n = [-a, a]^n$$

Integral of the n-volume of C^n:

$$I_0 = \int_{C^n} x^0 dx = (2a)^n$$

Integral of the square of any variable of C^n:

$$I_2 = \int_{C^n} x^2 dx = \int_{C^{n-1}} dx \int_{-\mathrm{a}}^{\mathrm{a}} x^2 dx = (2a)^{n-1} \int_{-\mathrm{a}}^{\mathrm{a}} x^2 dx = (2a)^{n-1} \left[\frac{1}{3} x^3\right]_{-a}^{a}$$
$$= (2a)^{n-1} \left(\frac{2}{3} a^3\right) = \frac{2^n}{3} a^{n+2}$$

Radius of the n-sphere on which the vertices of the Q^n must lie on:

$$r = \sqrt{n \frac{I_2}{I_0}} = \sqrt{n \left({}^{\frac{2^n}{3} a^{n+2}} \Big/ {}_{2^n a^n} \right)} = \sqrt{n \frac{a^2}{3}} = a\sqrt{\frac{n}{3}}$$

Then, for $n > 3$:

$$r > a$$

Source: own development based on a personal communication with Zimmermann (2010)

In order to avoid using points outside C^n, Stroud designed his formula to generate points which are rotated in order to take advantage of the distribution of space in C^n. With this design, he succeeded in presenting a formula whose points are inside C^n for all n.

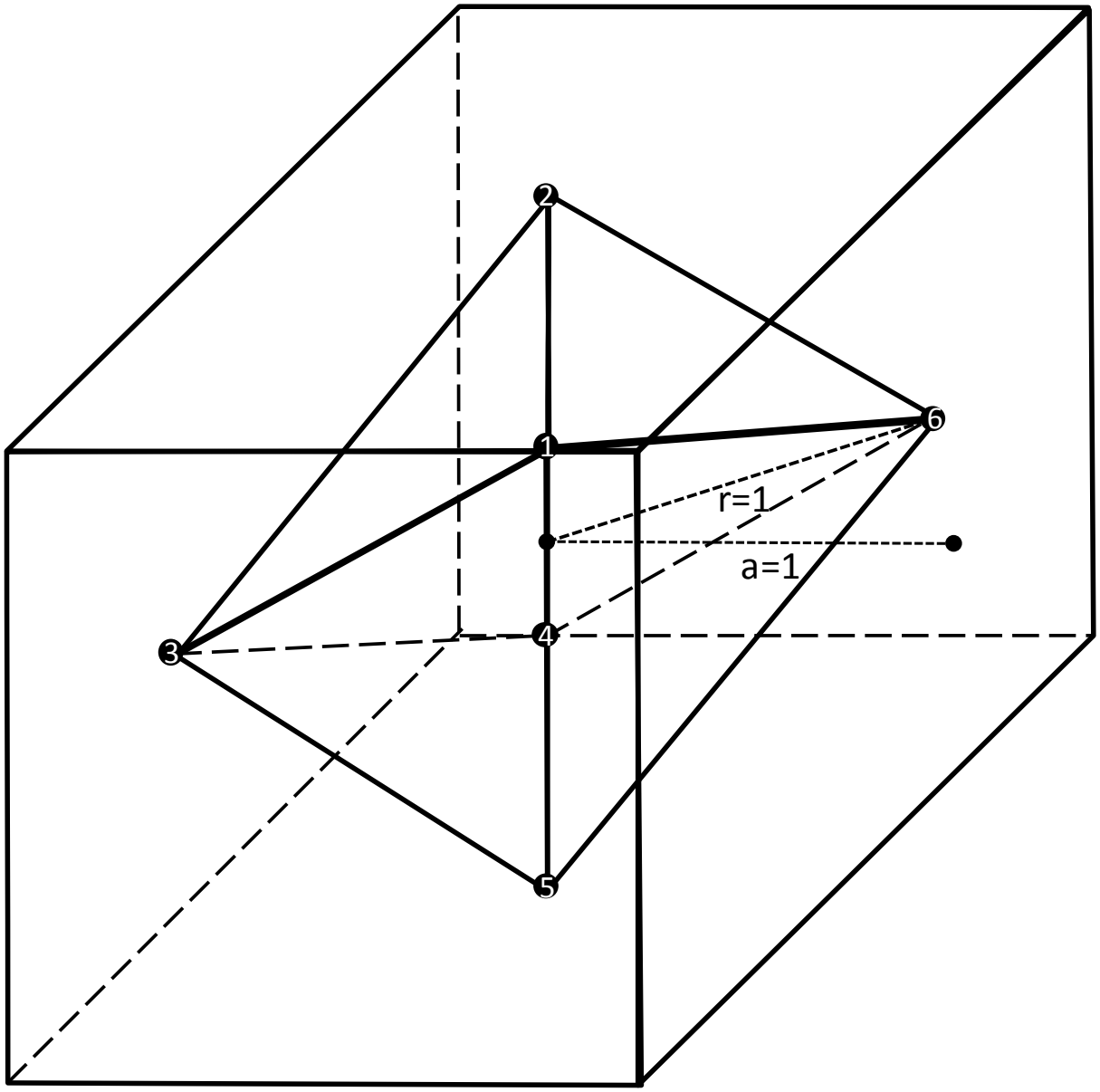
Figure 3.1 shows the quadrature points obtained with Stroud's formula for $n = 3$. Note that the vertices are enumerated – where vertices 1-4, 2-5, and 3-6 are opposed to each other – and that the point in the middle indicates the center of the region. It can be observed how the points are the vertices of a Q^n whose centroid coincides with the centroid of the region - C^n with vertices $(\pm 1, \pm 1, \cdots, \pm 1)$ - and lie on an n-sphere of radius $\mathrm{r} = \mathrm{a}\sqrt{\mathrm{n}/3} = 1$ for that C^{n}. Also, it can be seen how the Q^{n} is rotated, which exemplifies how the points are interior to C^{n} for higher dimensions.

Finally, remember that the formula produces quadratures of equal weights, which are given by (adapted from Stroud, 1957):

$$w_k = \frac{1}{2n} I_0 = \frac{(2a)^n}{2n} \qquad \text{for all } k = 1,2,\cdots,N, \tag{3.1}$$

where the n-volume of the region, I_0, is divided by the number of quadrature points, $N = 2n$. Then, by the weighted sum of the function values at the specified points, $f(\boldsymbol{\gamma}_k)$ for all $k = 1,2,\cdots,N$, one obtains an approximation of the $(n+1)$-volume under the graph of the multivariate function.

Figure 3.1: Visualization of the quadrature points generated with Stroud's formula for the n-cube with vertices $(\pm 1, \pm 1, \ldots, \pm 1)$, for $n = 3$



Matrix $\boldsymbol{\Gamma}$ with $n = 3$ variables and $N = 6$ quadrature points:

	$\boldsymbol{\gamma}_1$	$\boldsymbol{\gamma}_2$	$\boldsymbol{\gamma}_3$	$\boldsymbol{\gamma}_4$	$\boldsymbol{\gamma}_5$	$\boldsymbol{\gamma}_6$
$\boldsymbol{\gamma}_1$	0.41	-0.41	-0.82	-0.41	0.41	0.82
$\boldsymbol{\gamma}_2$	0.71	0.71	0.00	-0.71	-0.71	0.00
$\boldsymbol{\gamma}_3$	-0.58	0.58	-0.58	0.58	-0.58	0.58

Source: own development

The Formula in Detail

The formula has 2 parameters, the variables – all i – and the quadratures – all k. The expressions $2r-1$ and $2r$ indicate the odd and the even variables, thus, γ_{2r-1} and γ_{2r} define the coordinates of the k^{th} quadrature point as r increases, for problems with even n. In the case of odd n, the last coordinate, γ_n, is given by $(-1)^k/\sqrt{3}$. In this way, the matrix:

$$\mathbf{\Gamma} = \begin{bmatrix} \gamma_{1,1} & \cdots & \gamma_{1,N} \\ \vdots & \ddots & \vdots \\ \gamma_{n,1} & \cdots & \gamma_{n,N} \end{bmatrix} \tag{3.2}$$

of N quadrature points is computed.

The example from Figure 3.1 helps to understand how the different components of the formula work. The following is an analysis of the generation of those points.

The multivariate integration problem is of 3 variables – thus, 6 quadratures. Now, from the matrix $\mathbf{\Gamma}$ focus first on the approximation of $\boldsymbol{\gamma}_1$ - the first variable or row. In this case, it is necessary to use γ_{2r-1} 6 times as k gradually increases and all other terms, in γ_{2r-1}, remain constant. If we set $\alpha = (2r-1)k/n$, then, with increasing k we obtain a constantly increasing value of $\alpha\pi$. Now, by computing $\cos\alpha\pi$ we get values which go from -1 to 1 and would repeat once $\alpha\pi > 2\pi$ - this occurs in problems with higher n. Moreover, α is formulated in such a way that every value obtained with $\cos\alpha\pi$ has a negative or positive counterpart or is zero. This behavior is obtained due to the characteristic wave shape of the cosines function and to the formulation of α. Finally, the values obtained with $\cos\alpha\pi$ are multiplied with the factor $\sqrt{2/3}$, so that the symmetric structure observed before is transmitted to the 6 values in $\boldsymbol{\gamma}_1$; consequently $E[\boldsymbol{\gamma}_1] = 0$. For the approximation of $\boldsymbol{\gamma}_2$ the procedure is repeated but with the corresponding sine function, which does not alter the symmetric behavior observed for $\boldsymbol{\gamma}_1$, so that $E[\boldsymbol{\gamma}_2] = 0$. For $\boldsymbol{\gamma}_3$, the values of $(-1)^k/\sqrt{3}$ change between $1/\sqrt{3}$ and $-1/\sqrt{3}$ as k takes even or odd values respectively, so that $E[\boldsymbol{\gamma}_3] = 0$ too.

Now, note that the coordinates, $\gamma_{i,k}$ for all i and all k, take minimum and maximum values of $\pm\sqrt{2/3}$, which occur when $\cos\alpha\pi$ or $\sin\alpha\pi$, in γ_{2r-1} and γ_{2r}, take the values of ± 1. These minimum and maximum distances are obtained from (own development):

$$\min[\boldsymbol{\gamma}], \max[\boldsymbol{\gamma}] = \left(\pm \sqrt{\int_0^1 x^2 dx} \right) \sqrt{2} = \left(\pm \sqrt{\left[\frac{1}{3} x^3 \right]_0^1} \right) \sqrt{2} = \left(\pm \sqrt{\frac{1}{3}} \right) \sqrt{2} = \pm \sqrt{\frac{2}{3}}. \tag{3.3}$$

Note that the expression $\sqrt{\int_0^1 x^2 dx}$ is the square root of the integral of the square of any variable of C^n in the interval $[0,1]$, where $I_0 = 1$.

3.1.2 Adaptations Required for the Multivariate Standard Normal Distribution

From 2.1.2, we know that we can make a correspondence between the region of integration of numerical integration formulas and the multivariate PDF of discrete approximations of probability distributions. Therefore, we can say that Arndt and Artavia et al. adapted Stroud's quadrature points for symmetric regions to build a formula for the multivariate standard normal PDF.

In this case, the size of the Q^n changes as the region of integration changes. This size is captured by the magnitude of the radius, r, of the n-sphere on which the vertices of the Q^n must lie on (see Box 3.3).

Box 3.3: The radius of the n-sphere on which the vertices of the n-octahedron for the multivariate standard normal PDF must lie on

Region of integration:

The multivariate standard normal PDF, $\mathbf{x} \sim N(\mathbf{0}, \mathbf{I}_n)$

Integral of the measure – the n-volume – of the multivariate standard normal PDF:

$$I_0 = \int_{\mathbb{R}^n} x^0 \frac{1}{(2\pi)^{n/2}} e^{-\frac{\|\mathbf{x}\|^2}{2}} d\mathbf{x} = (1)^n = 1$$

Integral of the square of any variable of the multivariate standard normal PDF over $\mathbb{R}^n$:

$$I_2 = \int_{\mathbb{R}^n} x_i^2 \frac{1}{(2\pi)^{n/2}} e^{-\frac{\|\mathbf{x}\|^2}{2}} d\mathbf{x}$$

$$= \int_{\mathbb{R}^{n-1}} \frac{1}{(2\pi)^{(n-1)/2}} e^{-\frac{\|\mathbf{x}\|^2 - x_i^2}{2}} d\mathbf{x} \int_{-\infty}^{\infty} x_i^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{x_i^2}{2}} dx$$

$$= (1)^{n-1} 1 = 1$$

Radius of the n-sphere on which the vertices of the n-octahedron will lie on:

$$r = \sqrt{n \frac{I_2}{I_0}} = \sqrt{n}$$

Then, of course, for any n, r is interior to E^n.

Source: own formulation based on a personal communication with Zimmermann (2010)

Note that the multivariate standard normal PDF presents no restrictions of space and that consequently, the quadrature points – and with them, the n-sphere on which these must lie on – are inside the region of integration for any n, no matter the rotation that the Q^n presents.

The weights of the Stroud points are given by $w_k = \frac{I_0}{2n}$ (Stroud, 1957); thus, applied to the multivariate standard normal PDF we obtain:

$$p_k = \frac{1}{2n} I_0 = \frac{1}{2n} \qquad \text{for all } k = 1,2,\cdots,N. \tag{3.4}$$

Note that in this case, since we are working with the approximation of probability distributions, the weights from Stroud are associated with probabilities (for more details on the correspondence between numerical integration formulas and discrete approximations of PDFs see Section 2.1.2).

3.1.3 Arndt's Proposition

Arndt proposed to uses Stroud's formula in its conception for the C^n only adapting the size of the Q^n for the multivariate standard normal distribution as presented in Box 3.3. Arndt's formula is summarized in Box 3.4.

Box 3.4: Arndt's formula for the multivariate standard normal PDF

Let $\boldsymbol{\gamma}_k$ denote the point $(\gamma_1, \gamma_2, \cdots, \gamma_n)^T$, where

$$\gamma_{2r-1} = \sqrt{2}\cos\frac{(2r-1)k\pi}{n} \qquad \gamma_{2r} = \sqrt{2}\sin\frac{(2r-1)k\pi}{n}$$

$$\text{for } r = 1,2,\cdots,\left[\frac{1}{2}n\right]$$

($\left[\frac{1}{2}n\right]$ is the greatest integer not exceeding $\frac{1}{2}n$), and if n is odd:

$$\gamma_n = (-1)^k.$$

Then, the quadrature points $\boldsymbol{\gamma}_1, \cdots, \boldsymbol{\gamma}_N$ for $N = 2n$, satisfy the conditions of Stroud's theorem and, of course, all are interior to the multivariate standard normal PDF.

Source: own formulation based on Arndt (1996, p. 11)

Even though it is not needed to avoid using points outside the multivariate standard normal PDF, Arndt uses Stroud's formula for the C^n, which rotates the Q^n. This can be observed in Figure 3.2, which shows the Q^n for $n = 3$ obtained using Arndt's proposition. It can be seen that the only difference of the Q^n for the multivariate standard normal PDF with respect to the Q^n for the C^n with

vertices $(\pm 1, \pm 1, \cdots, \pm 1)$ is that the radius, r, is larger (see Figure 3.1 for comparison).

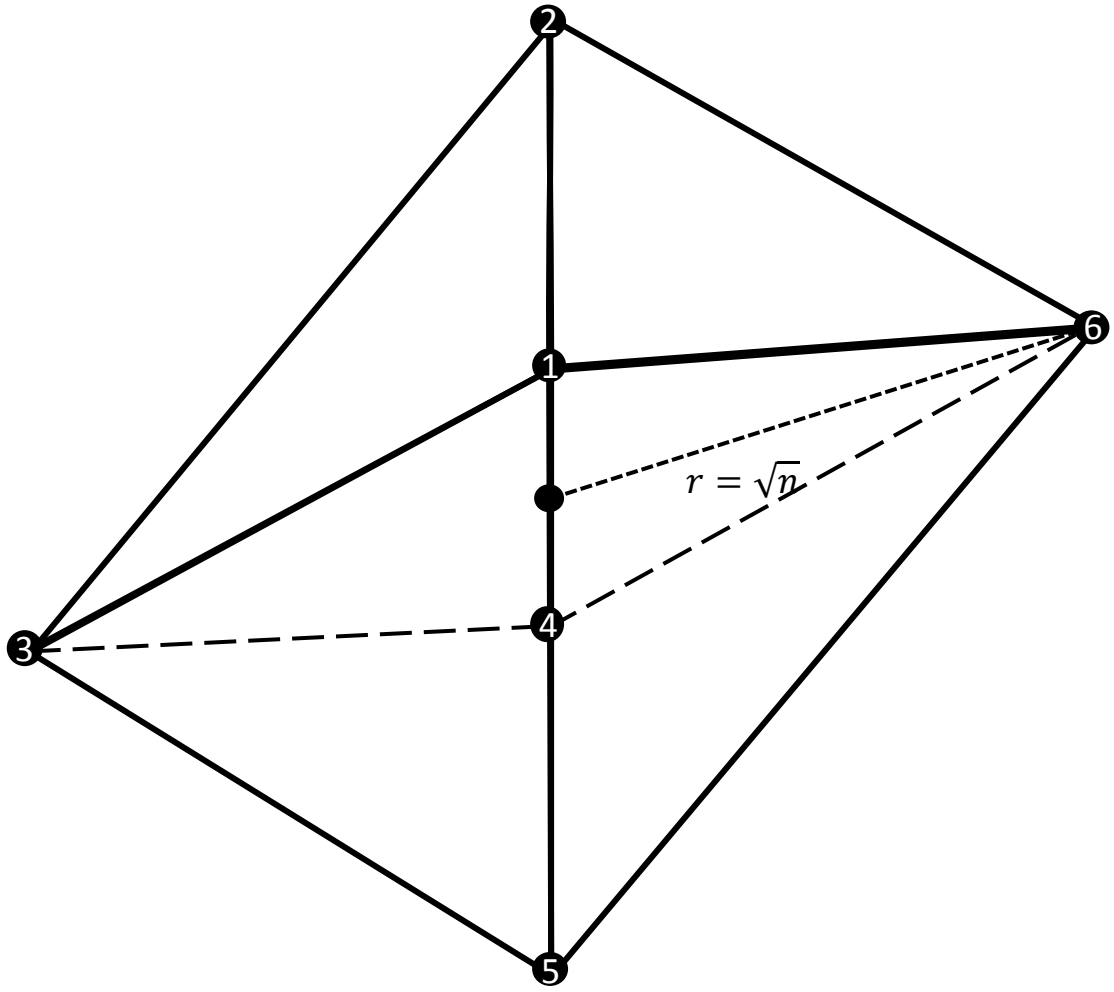
The Formula in Detail

The formula works in the same manner as Stroud's formula for the C^n, the only difference is that now the minimum and maximum values of discrete approximation of the PDFs of the stochastic variables are $\pm\sqrt{2}$. These minimum and maximum distances are obtained from:

$$\min[\boldsymbol{\gamma}], \max[\boldsymbol{\gamma}] = \left(\pm \sqrt{\int_{-\infty}^{\infty} x^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx} \right) \sqrt{2} = \pm\sqrt{2}. \tag{3.5}$$

Note that, in correspondence with (3.3), the expression $\sqrt{\int_{-\infty}^{\infty} x^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx}$ is the square root of the integral of the square of any variable times its probability density in the interval $[-\infty, \infty]$, where $I_0 = 1$. This expression takes the value of 1; thus, the minimum and maximum values in the discrete approximation of any variable are $\pm\sqrt{2}$. In this way, points lying on the tails of the standard normal distribution are avoided.

Figure 3.2: Visualization of the quadrature points generated with Arndt's proposition for the multivariate standard normal PDF, for $n = 3$



Matrix $\boldsymbol{\Gamma}$ with $n = 3$ variables and $N = 6$ quadrature points:

	$\boldsymbol{\gamma}_1$	$\boldsymbol{\gamma}_2$	$\boldsymbol{\gamma}_3$	$\boldsymbol{\gamma}_4$	$\boldsymbol{\gamma}_5$	$\boldsymbol{\gamma}_6$
$\boldsymbol{\gamma}_1$	0.71	-0.71	-1.41	-0.71	0.71	1.41
$\boldsymbol{\gamma}_2$	1.22	1.22	0.00	-1.22	-1.22	0.00
$\boldsymbol{\gamma}_3$	-1.00	1.00	-1.00	1.00	-1.00	1.00

Source: own development

3.1.4 Artavia et al.'s Proposition

Artavia et al. (2009) use Stroud's theorem – not the formula – and propose the Q^n with vertices on the coordinate axes as the integration formula (see Box 3.5).

Box 3.5: Artavia et al.'s formula for the multivariate standard normal PDFs

Let $\boldsymbol{\gamma}_k$ denote the point $(\gamma_1, \gamma_2, \cdots, \gamma_n)^T$, where

$$\boldsymbol{\gamma}_1 = \left(\sqrt{n}, 0, \cdots, 0\right)^T \qquad \boldsymbol{\gamma}_2 = \left(-\sqrt{n}, 0, \cdots, 0\right)^T$$
$$\boldsymbol{\gamma}_3 = \left(0, \sqrt{n}, \cdots, 0\right)^T \qquad \boldsymbol{\gamma}_4 = \left(0, -\sqrt{n}, \cdots, 0\right)^T$$
$$\vdots \qquad\qquad \vdots$$
$$\boldsymbol{\gamma}_{N-1} = \left(0, 0, \cdots, \sqrt{n}\right)^T \qquad \boldsymbol{\gamma}_N = \left(0, 0, \cdots, -\sqrt{n}\right)^T.$$

Then, the quadrature points $\boldsymbol{\gamma}_1, \cdots, \boldsymbol{\gamma}_N$ for $N = 2n$, satisfy the conditions of Stroud's theorem and, of course, all are interior to the multivariate standard normal PDF

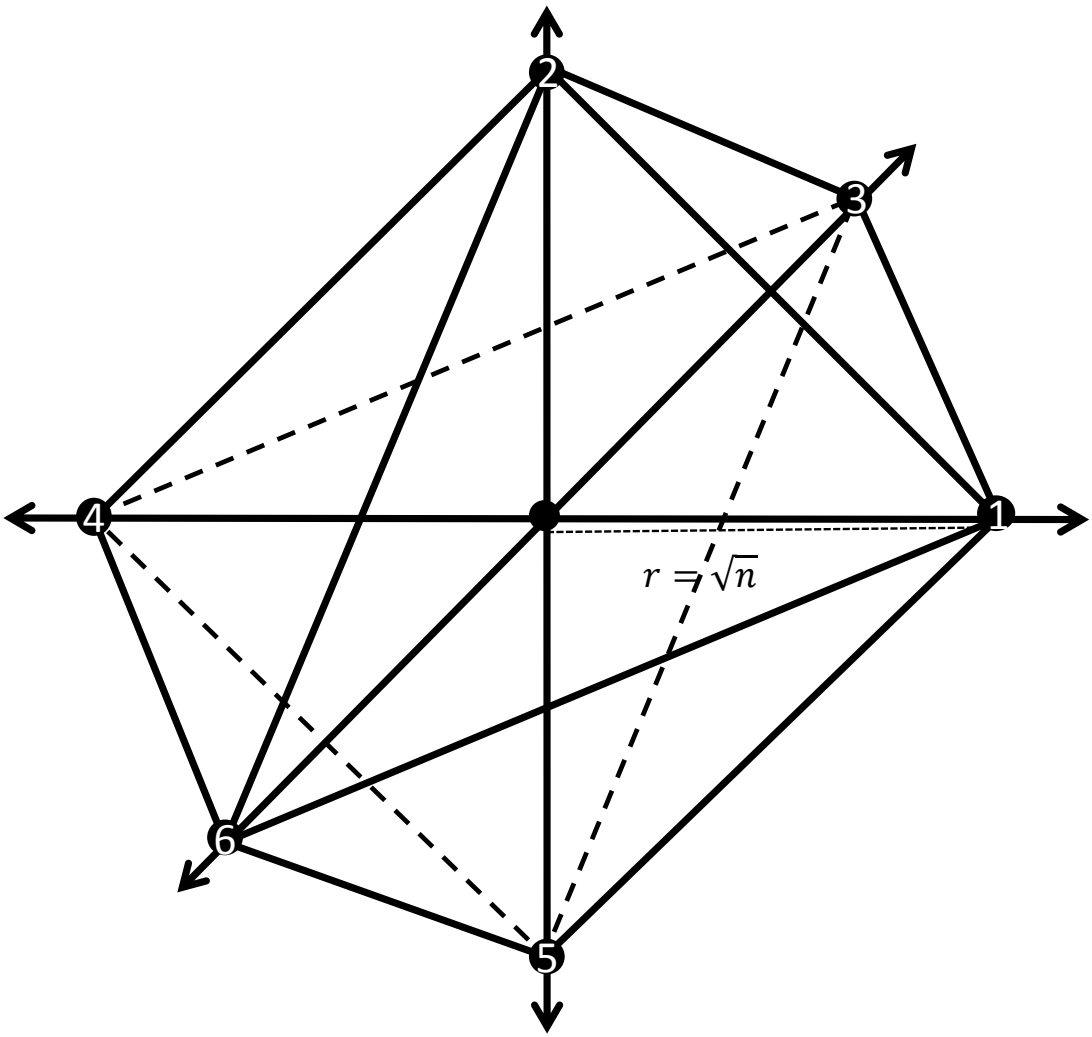
Source: own formulation based on Artavia et al. (2009)

As already mentioned, for the multivariate standard normal PDF, the quadrature points are always inside the region of integration for any n, no matter the rotation of the Q^n (see Box 3.3). Thus, Artavia et al. proposed no rotation (see Figure 3.3).

The Formula in Detail

Since the vertices of the Q^n lie on the coordinate axes, the unnecessary calculations in order to give the Q^n the desired rotation required by Stroud in order to get the points inside the C^n for all n are avoided and the formula becomes extremely simple. It is reduced to the computation of $\pm\sqrt{n}$. In this way, the minimum and maximum values of the discrete approximation of the PDFs of the stochastic variables are given by $\pm\sqrt{n}$. Note that for integration problems of low n, the points will lie near to the standard deviation of the approximated PDF; for example, for $\sigma = 1$ and $n = 3$, $\sqrt{3} \approx 1.73$. On the other hand, for integration problems of large n, the points will be progressively lying further from the mean; for example, for $\sigma = 1$ and $n = 42$ (as in the computation of the quadratures for the multivariate standard normal PDF – quadratures before transformation – in the stochastic version of ESIM), $\pm\sqrt{42} \approx \pm 6.48$. In this way, with large n, this formula generates minimum and maximum points which are very unlikely to occur.

Figure 3.3: Visualization of the quadrature points generated with Artavia et al.'s proposition for the multivariate standard normal PDF, for $n = 3$



Matrix $\boldsymbol{\Gamma}$ with $n = 3$ variables and $N = 6$ quadrature points:

	$\boldsymbol{\gamma}_1$	$\boldsymbol{\gamma}_2$	$\boldsymbol{\gamma}_3$	$\boldsymbol{\gamma}_4$	$\boldsymbol{\gamma}_5$	$\boldsymbol{\gamma}_6$
$\boldsymbol{\gamma}_1$	1.73	0.00	0.00	-1.73	0.00	0.00
$\boldsymbol{\gamma}_2$	0.00	1.73	0.00	0.00	-1.73	0.00
$\boldsymbol{\gamma}_3$	0.00	0.00	1.73	0.00	0.00	-1.73

Source: own development

3.1.5 Moments of the Discrete Approximations

Finally, for the quadratures produced with the formulas from Arndt and Artavia et al. for the multivariate standard normal PDF, the following central moments up to degree three, $|\mathbf{r}| = 3$, are obtained (adapted from Artavia et al., 2009):

(3.6)
$$|\mathbf{r}| = 0: \quad \frac{1}{2n}\sum_{k=1}^{2n} 1 = 1,$$

$$|\mathbf{r}| = 1: \quad \frac{1}{2n}\sum_{k=1}^{2n} \gamma_{k,i} = 0 \qquad \text{for all } i = 1,2,\cdots,n$$

$$\Leftrightarrow \frac{1}{2n}\sum_{k=1}^{2n} \boldsymbol{\gamma}_k = \mathbf{0},$$

$$|\mathbf{r}| = 2: \quad \frac{1}{2n}\sum_{k=1}^{2n} \gamma_{k,i_1}\gamma_{k,i_2} = \sigma_{\mathrm{i}_1,\mathrm{i}_2} \quad \text{for all } i_1, i_2 = 1,2,\cdots,n$$

$$\Leftrightarrow \frac{1}{2n}\sum_{k=1}^{2n} \boldsymbol{\gamma}_k\boldsymbol{\gamma}_k^T = \mathbf{I}_n,$$

$$|\mathbf{r}| = 3: \quad \frac{1}{2n}\sum_{k=1}^{2n} \gamma_{k,i_1}\gamma_{k,i_2}\gamma_{k,i_3} = 0$$

$$\text{for all } i_1, i_2, i_3 = 1,2,\cdots,n.$$

The moments from above are shown in two ways, first for the marginal PDFs – the PDF of every stochastic variable – and then for the multivariate PDF as a whole. It can be seen that the moments until $|\mathbf{r}| = 3$ of the multivariate standard normal distribution are matched. Thus, we indeed have a quadrature formula of the desired degree 3 of precision.

3.2 Transforming the Multivariate Standard Normal Quadratures

3.2.1 Inducing a Desired Covariance Matrix

The fact that in the last section I_2 was independent of the variable chosen stems from the symmetry of the regions C^n and $\mathbf{x} \sim N(\mathbf{0}, \mathbf{I}_n)$. However, in stochastic

modeling the integration regions are often asymmetric, presenting, for example, variables with variances of different sizes or with an important level of correlation. Thus, in this and the next section, it is studied how to get quadratures for the multivariate normal PDF, $\mathbf{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

'Efficient Quadratures' and Correlation: Methods Found in the Literature

Preckel and DeVuyst (1992) and DeVuyst (1993) presented the same method to induce a desired correlation to quadratures which can be summarized as follows:

- First, transform the empirical data, $\mathbf{Z}$, in order to get a new matrix $\mathbf{C}$ of independent stochastic variables. To do this they factorize the covariance matrix, using the diagonalization method (see Box 3.6), in the form $\boldsymbol{\Sigma}[\boldsymbol{z}] = \mathbf{U}\mathbf{D}\mathbf{U}^{\mathrm{T}}$ and compute $\mathbf{C} = \mathbf{U}^{\mathrm{T}}\mathbf{Z}$.
- Second, calculate a matrix of quadratures $\mathbf{Q}_{n \times N}$ using univariate quadrature formulas for the marginal distributions of the variables in $\mathbf{C}$.
- And finally, use the inverse transformation $\mathbf{Q}^{*} = \mathbf{U}\mathbf{Q}$ in order to get the matrix $\mathbf{Q}^{*}_{\mathbf{n} \times \mathbf{N}}$ of N quadratures with the desired $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.

This procedure works, nonetheless, they named the factorization $\boldsymbol{\Sigma}[\boldsymbol{z}] = \mathbf{U}\mathbf{D}\mathbf{U}^{\mathrm{T}}$ a Cholesky decomposition which could lead researchers in the wrong direction, e.g., in intending to carry on a decomposition as $\boldsymbol{\Sigma}[\boldsymbol{z}] = \mathbf{L}\mathbf{D}\mathbf{L}^{\mathrm{T}}$ (see Box 3.6). Furthermore, one must be aware that $\mathbf{U}$ must have the eigenvectors written in columns. It is important to identify this clearly in order to avoid calculation errors. For example, Preckel and DeVuyst suggested the decomposition $\boldsymbol{\Sigma}[\boldsymbol{z}] = \mathbf{U}^{\mathbf{T}}\mathbf{D}\mathbf{U}$ and were indicating the diagonalization method, but in their matrix $\mathbf{U}$ the vectors were in the rows. In this way their factorization is equal to that described in Box 3.6.

Arndt (1996) stated that the decomposition of the covariance matrix of the data could be obtained via Cholesky factorization using the following notation: $\boldsymbol{\Sigma}[\boldsymbol{z}] = \mathbf{L}\mathbf{D}\mathbf{L}^{\mathrm{T}}$ and indicated that $\boldsymbol{\Sigma}[\boldsymbol{z}]$ could be induced to the Q^{n} for the multivariate standard normal PDF from last section by computing $\boldsymbol{\Gamma}^{*} = \boldsymbol{\Gamma}\mathbf{L}\sqrt{\mathbf{D}}$.

There are two possible interpretations for that, a Cholesky decomposition in that form or the diagonalization method (see for both methods Box 3.6). Then, $\mathbf{L}$ and $\mathbf{D}$ are either the matrices from the Cholesky decomposition $\mathbf{L}\mathbf{D}\mathbf{L}^{\mathrm{T}}$ or the matrices $\mathbf{U}$ and $\mathbf{D}$ from the diagonalization method. Anyhow, in both cases $\boldsymbol{\Gamma}^{*} = \boldsymbol{\Gamma}\mathbf{L}\sqrt{\mathbf{D}}$ does not lead to the desired result. If we use the Cholesky factors, then the resulting quadratures do not have the desired $\boldsymbol{\Sigma}[\boldsymbol{z}]$. If, conversely, we use the factors from the diagonalization method, then the resulting quadratures are interestingly still independent and present a variance equal to the eigenvalues $\mathbf{D}$ of $\boldsymbol{\Sigma}[\boldsymbol{z}]$, but do not have the desired variance and covariance. This occurs

because the step $\mathbf{\Gamma U}$ only results in the rotation of the vectors in $\mathbf{\Gamma}$, but these remain orthogonal to each other with $\mathbf{\Sigma}[\boldsymbol{\gamma}] = \mathbf{I}_n$. The following step $\mathbf{\Gamma U}\sqrt{\mathbf{D}}$ means the extension (as opposed to compression) of the discrete approximation of the multivariate standard normal PDF.

Probably, with his statement, Arndt (1996) referred to the decomposition of $\mathbf{\Sigma}[\boldsymbol{z}]$ using the diagonalization method and to induce the desired variance and covariance by computing $\mathbf{\Gamma}^* = \mathbf{U}\sqrt{\mathbf{D}}\mathbf{\Gamma}$ as it is explained in the next subsection.

Preckel and DeVuyst (1992), DeVuyst (1993) and Arndt (1996) all presented practicable procedures to induce a desired covariance matrix to quadratures; however, it has not been clearly indicated how to decompose $\mathbf{\Sigma}[\boldsymbol{z}]$, with the Cholesky decomposition or with the diagonalization method. Artavia et al. (2009) elaborate a theoretical demonstration on how to induce correlation and by presenting different ways they clarify this issue. Currently, Preckel et al. (2010) and Horridge and Pearson (2011) have also made some contributions to the topic motivated on the developments exposed by Artavia et al. (2009). The expositions made by Artavia et al. (2009) are displayed in the next subsection.

'Efficient Quadratures' and Correlation: A Theoretical Demonstration

In this subsection, the following topic is treated: how to induce a desired covariance matrix, $\mathbf{\Sigma}[\boldsymbol{z}]$, to $\mathbf{\Gamma}$ with $\mathbf{\Sigma}[\boldsymbol{\gamma}] = \mathbf{I}_n$ in order to get new quadratures, $\mathbf{X}$, with $\mathbf{\Sigma}[\boldsymbol{x}] = \mathbf{\Sigma}[\boldsymbol{z}]$. With it, it is demonstrated how a linear transformation applied to some quadratures influences the covariance matrix of the approximation points. This demonstration leads to the different methods to induce the desired properties.

To begin with, we consider an equidistribution of N arbitrary points $\mathbf{x}_1, \cdots, \mathbf{x}_N \in \mathbb{R}^n$ with weight $1/N$ and mean $\mathrm{E}[\boldsymbol{x}] = (1/N)(\mathbf{x}_1 + \cdots + \mathbf{x}_\mathrm{N}) = \mathbf{0}$. In this case, the covariance matrix can be determined by gathering these points in a $n \times N$-matrix:

$$\boldsymbol{X} = \begin{bmatrix} x_{1,1} & \cdots & x_{1,N} \\ \vdots & \ddots & \vdots \\ x_{n,1} & \cdots & x_{n,N} \end{bmatrix} \tag{3.7}$$

and computing:

$$\mathbf{\Sigma}[\boldsymbol{x}] = \frac{1}{N}\mathbf{X}\mathbf{X}^\mathrm{T}. \tag{3.8}$$

For the vertices $\boldsymbol{\gamma}_1, \cdots, \boldsymbol{\gamma}_N$ of the Q^n for the multivariate standard normal PDF obtained with the formulas from Arndt and Artavia et al., this yields:

$$\mathbf{\Sigma}[\boldsymbol{\gamma}] = \frac{1}{N}\mathbf{\Gamma}\mathbf{\Gamma}^{\mathrm{T}} = \mathbf{I}_{\mathrm{n}} \tag{3.9}$$

as claimed before.

Now, let $\mathbf{A}$ be any regular $n \times n$-matrix and consider the points $\mathbf{x}_k = \mathbf{A}\boldsymbol{\gamma}_k$ for $k = 1,2,\cdots,N$, where $N = 2n$. This yields:

$$\mathbf{X} = [\mathbf{A}\boldsymbol{\gamma}_1|\cdots|\boldsymbol{A}\boldsymbol{\gamma}_N] = \mathbf{A}\mathbf{\Gamma} \tag{3.10}$$

with:

$$\mathrm{E}[\boldsymbol{x}] = \frac{1}{2n}\sum_{k=1}^{2n}\mathbf{x}_k = \frac{1}{2n}\mathbf{A}\sum_{k=1}^{2n}\boldsymbol{\gamma}_k = \mathbf{A}\,\mathrm{E}[\boldsymbol{\gamma}] = \mathbf{A}\,\mathbf{0} = \mathbf{0} \tag{3.11}$$

and:

$$\begin{aligned}\mathbf{\Sigma}[\boldsymbol{x}] = \frac{1}{2n}\mathbf{X}\mathbf{X}^{\mathrm{T}} = \frac{1}{2n}\mathbf{A}\mathbf{\Gamma}(\mathbf{A}\mathbf{\Gamma})^{\mathrm{T}} = \frac{1}{2n}\mathbf{A}\mathbf{\Gamma}\mathbf{\Gamma}^{\mathbf{T}}\mathbf{A}^{\mathbf{T}} = \mathbf{A}\frac{1}{2n}\mathbf{\Gamma}\mathbf{\Gamma}^{\mathbf{T}}\mathbf{A}^{\mathbf{T}} \\ = \mathbf{A}\mathbf{I}_n\mathbf{A}^{\mathrm{T}} = \mathbf{A}\mathbf{A}^{\mathrm{T}} = \mathbf{\Sigma}[\boldsymbol{z}].\end{aligned} \tag{3.12}$$

Thus, our problem is reduced to expressing the desired covariance matrix, $\mathbf{\Sigma}[\boldsymbol{z}]$, in the form $\mathbf{A}\mathbf{A}^{\mathrm{T}}$ for a regular square matrix $\mathbf{A}$. There are countless possibilities of doing this; Box 3.6 describes three standard methods.

Box 3.6 Three standard methods of decomposing the covariance matrix which can be used to induce first and second moments to the quadratures

1. Diagonalization (Principal Axes Transformation):

Since $\boldsymbol{\Sigma}[\boldsymbol{z}]$ is positive semidefinite, it can be written in the form $\boldsymbol{\Sigma}[\boldsymbol{z}] = \mathbf{U}\mathbf{D}\mathbf{U}^{\mathrm{T}}$, where $\mathbf{D}$ is the (non-negative) diagonal matrix of eigenvalues of $\boldsymbol{\Sigma}[\boldsymbol{z}]$ and $\mathbf{U}$ is orthogonal (consisting of the eigenvectors of $\boldsymbol{\Sigma}[\boldsymbol{z}]$). Notice that the vectors (eigenvectors) are the columns of $\mathbf{U}$. Then, letting $\mathbf{A} = \mathbf{U}\sqrt{\mathbf{D}}$ yields $\mathbf{A}\mathbf{A}^{\mathrm{T}} = \boldsymbol{\Sigma}[\boldsymbol{z}]$ as desired.

2. Cholesky decomposition:

The positive semidefinite matrix $\boldsymbol{\Sigma}[\boldsymbol{z}]$ has a Cholesky decomposition $\boldsymbol{\Sigma}[\boldsymbol{z}] = \mathbf{L}\mathbf{L}^{\mathrm{T}}$ where $\mathbf{L}$ is a lower triangular matrix as follows:

$$\mathbf{L} = \begin{bmatrix} L_{1,1} & 0 & \cdots & 0 \\ L_{2,1} & L_{2,2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ L_{n,1} & \cdots & L_{n,n-1} & L_{n,n} \end{bmatrix},$$

then, choose $\mathbf{A} = \mathbf{L}$.

The Cholesky decomposition can also be of the form $\boldsymbol{\Sigma}[\boldsymbol{z}] = \mathbf{L}\mathbf{D}\mathbf{L}^{\mathrm{T}}$ where $\mathbf{L}$ is a lower triangular matrix and $\mathbf{D}$ a diagonal matrix as follows:

$$\mathbf{L}\mathbf{D}\mathbf{L}^{\mathrm{T}} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ L_{2,1} & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ L_{n,1} & \cdots & L_{n,n-1} & 1 \end{bmatrix} \begin{bmatrix} D_1 & 0 & \cdots & 0 \\ 0 & D_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & D_{\mathrm{n}} \end{bmatrix} \begin{bmatrix} 1 & L_{1,2} & \cdots & L_{1,n} \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & L_{n-1,n} \\ 0 & \cdots & 0 & 1 \end{bmatrix},$$

then, choose $\mathbf{A} = \mathbf{L}\sqrt{\mathbf{D}}$.

3. Reverse Cholesky decomposition:

The positive semidefinite matrix $\boldsymbol{\Sigma}[\boldsymbol{z}]$ can also be decomposed in $\boldsymbol{\Sigma}[\boldsymbol{z}] = \mathbf{R}\mathbf{R}^{\mathrm{T}}$ where $\mathbf{R}$ is an upper triangular matrix. In this case, choose $\mathbf{A} = \mathbf{R}$ to obtain the desired conditions.

Source: own development based on Artavia et al. (2009, pp. 12-13)

It is worth noting that different factorizations $\boldsymbol{\Sigma}[\boldsymbol{z}] = \mathbf{A}\mathbf{A}^{\mathrm{T}} = \mathbf{B}\mathbf{B}^{\mathrm{T}}$ simply differ by an orthogonal matrix factor, e.g., $\mathbf{B} = \mathbf{A}\mathbf{O}$ for an orthogonal matrix $\mathbf{O}$; and consequently, each such matrix $\mathbf{O}$ will yield a different factorization. Therefore, choosing a different factorization, e.g., $\mathbf{B}$ instead of $\mathbf{A}$, will result in $\mathbf{B}\boldsymbol{\Gamma} = \mathbf{A}\mathbf{O}\boldsymbol{\Gamma}$. This simply means rotating the quadratures for the multivariate standard normal distribution before applying the transformation to induce the desired covariance.

In summary, for each choice of $\mathbf{A}$ with $\mathbf{A}\mathbf{A}^{\mathrm{T}} = \boldsymbol{\Sigma}[\boldsymbol{z}]$, the columns of the matrix $\mathbf{X} = \mathbf{A}\boldsymbol{\Gamma}$ yield quadrature points with $\mathrm{E}[\boldsymbol{x}] = \mathbf{0}$ (the zero vector) and $\boldsymbol{\Sigma}[\boldsymbol{x}] = \boldsymbol{\Sigma}[\boldsymbol{z}]$.

3.2.2 Inducing Desired Means

To additionally introduce nonzero means it suffices to add the vector $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_n)^{\mathrm{T}}$ to each of the transformed points from above. For example, if we let $\mathbf{Y} = (\mathbf{y}_1 | \cdots | \mathbf{y}_N)$ be the matrix of quadratures with $\boldsymbol{\mu} \neq \mathbf{0}$ then, $\mathbf{y}_k$ can be obtained by

$$\mathbf{y}_k = \mathbf{A}\boldsymbol{\gamma}_k + \boldsymbol{\mu}. \tag{3.7}$$

The transformations in Sections 3.2.1 and 3.2.2 generate a discrete approximation of the multivariate PDF with the desired first and second moments. That all third centered moments are still zero – as is the case with any normal distribution – is a consequence of the symmetry about the vector $\boldsymbol{\mu}$. Therefore, the constructed points yield a quadrature formula for multivariate normal distributions with degree of exactness 3.

3.2.3 Inducing Variances and Correlation Separately

In Section 3.2.1, we have studied how to get quadratures with some desired variances and covariance by using the factorization of the covariance matrix. Those procedures induce both variances and covariance at the same time when computing $\mathbf{A}\boldsymbol{\Gamma}$. In some circumstances it may be desirable to induce these two separately. For example, a possible research question may be to analyze the consequences of a probable increase of extreme weather conditions on the agricultural sector. For such a case the variances of stochastic variables would not be based on historical observations but on future predictions. Nonetheless, one may be interested in conserving the same correlation obtained from historical data. In this situation, it would be necessary to induce the observed correlation and the predicted variance in two separate steps.

We start by factoring the correlation matrix $\mathbf{P}[\boldsymbol{z}]$ in the form $\mathbf{A}\mathbf{A}^{\mathrm{T}}$. Then, by calculating $\mathbf{A}\boldsymbol{\Gamma}$ we get a new matrix $\mathbf{Y}$ (of vectors $\mathbf{y}_1, \cdots, \mathbf{y}_N$ and stochastic variables $\boldsymbol{y}_1, \cdots, \boldsymbol{y}_n$) with $\mathrm{E}[\boldsymbol{y}] = \mathbf{0}$ and $\boldsymbol{\Sigma}[\boldsymbol{y}] = \mathbf{P}[\boldsymbol{z}]$.

In the second step, we estimate $\mathbf{D}_\sigma\mathbf{Y}$ to get the final matrix $\mathbf{X}$, where $\mathbf{D}_\sigma$ is a diagonal matrix with the desired standard deviations σ_i in its diagonal:

$$\mathbf{D}_\sigma = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_\mathrm{n} \end{bmatrix}. \tag{3.13}$$

Altogether, this yields $\mathbf{X} = \mathbf{D}_\sigma \mathbf{A}\boldsymbol{\Gamma}$ with $\mathrm{E}[\boldsymbol{x}] = \mathbf{0}$ and $\boldsymbol{\Sigma}[\boldsymbol{x}] = \mathbf{D}_\sigma \mathbf{A}\mathbf{A}^\mathbf{T}\mathbf{D}_\sigma = \mathbf{D}_\sigma \mathbf{P}\mathbf{D}_\sigma$ as desired.

3.3 An Example of the How to Do It

Some authors assert that even though formulas to include covariance have been given, no detailed explanation has been provided yet (see e.g., Horridge and Pearson, 2011). This section intends to fill in this gap and seeks to contribute to the further determination of the qualities and differences between the quadrature formulas presented in this study.

Artavia et al. (2009) solve a small example in which quadrature points are generated using their proposed formula in combination with the Cholesky decomposition. The presentation provided here is broader, solving examples with different formulas for the approximation of $\mathbf{x} \sim N(\mathbf{0}, \mathbf{I}_n)$ (the approaches used are the LHS method, Arndt's formula, and Artavia et al.'s formula). Furthermore, the effect of different transformations to get formulas for the approximation of $\mathbf{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is exemplified (transformations via Cholesky decomposition or the diagonalization method; see box 3.6). Finally, the generated points are evaluated with respect to their descriptive statistics and probability distribution, and with respect to their accuracy of approximation of the first and second moments of $\mathbf{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

In the example below, the different quadratures are generated based on annual yield deviates from a trend: $\frac{yield}{trend} - 1$. These calculations are done for 6 variables, for 13 years which corresponds to data from three crops in two countries. The mean of every stochastic variables is assumed to be zero and the standard deviation ranges among 0.05 to 0.15 (5% to 15%). Boxes 3.7 to 3.11 show the deviates and the different steps required to obtain the quadratures. Boxes 3.12 to 3.14 show the accuracy of the generated quadratures.

Box 3.7: Example yield deviates* and their covariance and correlation matrices

Matrix of deviates for $n = 1,2,\cdots,6$ stochastic variables and $m = 1,2,\cdots,13$ observations:

$$\mathbf{Z}_{n\times m} = \begin{bmatrix} -0.074 & -0.085 & -0.005 & 0.008 & 0.116 & 0.040 & 0.103 & 0.001 & 0.113 & -0.101 & -0.130 & 0.064 & -0.051 \\ 0.171 & 0.102 & 0.099 & -0.348 & -0.109 & -0.007 & 0.187 & -0.058 & 0.040 & -0.083 & -0.173 & 0.150 & 0.056 \\ -0.045 & -0.034 & 0.002 & 0.032 & 0.046 & 0.032 & -0.028 & 0.040 & 0.067 & -0.054 & -0.148 & 0.095 & -0.006 \\ 0.092 & 0.048 & -0.075 & 0.030 & 0.036 & -0.068 & -0.122 & -0.079 & 0.007 & -0.034 & 0.017 & 0.116 & 0.032 \\ -0.078 & -0.003 & 0.098 & -0.065 & 0.034 & 0.012 & 0.177 & 0.038 & -0.090 & -0.161 & -0.132 & 0.062 & 0.108 \\ 0.019 & -0.057 & -0.038 & 0.029 & 0.069 & 0.026 & 0.007 & -0.004 & -0.006 & -0.040 & -0.071 & 0.041 & 0.025 \end{bmatrix}$$

Covariance $\mathbf{\Sigma}[\boldsymbol{z}]$ and correlation $\mathbf{P}[\boldsymbol{z}]$ matrices:

$$\mathbf{\Sigma}[\mathbf{z}] = \begin{bmatrix} 0.0065 & 0.0018 & 0.0038 & -0.0011 & 0.0038 & 0.0021 \\ 0.0018 & 0.0219 & 0.0009 & 0.0000 & 0.0072 & 0.0000 \\ 0.0038 & 0.0009 & 0.0037 & 0.0004 & 0.0021 & 0.0017 \\ -0.0011 & 0.0000 & 0.0004 & 0.0046 & -0.0021 & 0.0007 \\ 0.0038 & 0.0072 & 0.0021 & -0.0021 & 0.0093 & 0.0014 \\ 0.0021 & 0.0000 & 0.0017 & 0.0007 & 0.0014 & 0.0016 \end{bmatrix} \quad \mathbf{P}[\boldsymbol{z}] = \begin{bmatrix} 1 & 0.15 & 0.77 & -0.19 & 0.49 & 0.66 \\ 0.15 & 1 & 0.10 & 0.00 & 0.50 & 0.00 \\ 0.77 & 0.10 & 1 & 0.10 & 0.36 & 0.69 \\ -0.19 & 0.00 & 0.10 & 1 & -0.32 & 0.24 \\ 0.49 & 0.50 & 0.36 & -0.32 & 1 & 0.37 \\ 0.66 & 0.00 & 0.69 & 0.24 & 0.37 & 1 \end{bmatrix}$$

Source: own calculations *Yield deviates: $\frac{yield}{trend} - 1$

Box 3.8: Example decompositions of the covariance matrix

Cholesky decomposition, $\mathbf{\Sigma}[\boldsymbol{z}] = \mathbf{L}\mathbf{L}^{\mathrm{T}}$:

$$\mathbf{L} = \begin{bmatrix} 0.0809 & 0 & 0 & 0 & 0 & 0 \\ 0.0225 & 0.1462 & 0 & 0 & 0 & 0 \\ 0.0467 & -0.0009 & 0.0391 & 0 & 0 & 0 \\ -0.0130 & 0.0023 & 0.0259 & 0.0615 & 0 & 0 \\ 0.0469 & 0.0420 & -0.0005 & -0.0257 & 0.0683 & 0 \\ 0.0262 & -0.0038 & 0.0112 & 0.0117 & 0.0097 & 0.0227 \end{bmatrix} \quad \mathbf{L}^{\mathrm{T}} = \begin{bmatrix} 0.0809 & 0.0225 & 0.0467 & -0.0130 & 0.0469 & 0.0262 \\ 0 & 0.1462 & -0.0009 & 0.0023 & 0.0420 & -0.0038 \\ 0 & 0 & 0.0391 & 0.0259 & -0.0005 & 0.0112 \\ 0 & 0 & 0 & 0.0615 & -0.0257 & 0.0117 \\ 0 & 0 & 0 & 0 & 0.0683 & 0.0097 \\ 0 & 0 & 0 & 0 & 0 & 0.0227 \end{bmatrix}$$

Diagonalization method, $\mathbf{\Sigma}[\boldsymbol{z}] = \mathbf{U}\sqrt{\mathbf{D}}\left(\mathbf{U}\sqrt{\mathbf{D}}\right)^{\mathrm{T}}$ where $\mathbf{U}$ is the matrix of eigenvectors - in columns-, and $\mathbf{D}$ the matrix of eigenvalues:

$$\mathbf{U} = \begin{bmatrix} 0.1948 & -0.6219 & 0.2191 & -0.4294 & 0.5276 & -0.2539 \\ 0.8666 & 0.4257 & 0.1363 & -0.2094 & -0.0102 & 0.0729 \\ 0.1120 & -0.4124 & 0.3829 & -0.1099 & -0.8087 & -0.0683 \\ -0.0494 & 0.1762 & 0.7477 & 0.5307 & 0.2094 & -0.2864 \\ 0.4402 & -0.4107 & -0.4064 & 0.6718 & -0.0010 & -0.1453 \\ 0.0483 & -0.2496 & 0.2502 & 0.1636 & 0.1537 & 0.9069 \end{bmatrix} \quad \mathbf{D} = \begin{bmatrix} 0.0261 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.0115 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.0059 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.0030 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.0008 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.0004 \end{bmatrix}$$

$$\mathbf{U}\sqrt{\mathbf{D}} = \begin{bmatrix} 0.0315 & -0.0666 & 0.0168 & -0.0234 & 0.0150 & -0.0053 \\ 0.1399 & 0.0456 & 0.0104 & -0.0114 & -0.0003 & 0.0015 \\ 0.0181 & -0.0441 & 0.0293 & -0.0060 & -0.0230 & -0.0014 \\ -0.0080 & 0.0189 & 0.0572 & 0.0289 & 0.0060 & -0.0060 \\ 0.0711 & -0.0440 & -0.0311 & 0.0366 & 0.0000 & -0.0030 \\ 0.0078 & -0.0267 & 0.0192 & 0.0089 & 0.0044 & 0.0190 \end{bmatrix}$$

Source: own calculations

Box 3.9: Example LHS, Arndt, and Artavia et al. quadratures for the multivariate standard normal PDF for 6 stochastic variables

$$\boldsymbol{\Gamma}_{LHS} = \begin{bmatrix} -0.30 & -0.67 & -1.08 & -0.80 & 0.65 & 0.94 & 0.37 & 0.13 & 1.58 & -1.52 & 0.97 & -0.01 \\ -0.19 & 0.35 & -1.07 & 1.31 & -0.55 & 0.63 & -1.54 & 0.83 & -0.40 & 3.18 & -0.80 & 0.02 \\ 1.58 & 0.85 & -0.54 & 0.33 & -1.01 & -2.18 & 1.35 & 0.01 & -0.75 & -0.16 & 0.58 & -0.39 \\ -2.74 & -0.10 & 0.24 & 0.04 & 0.51 & -1.03 & 1.05 & 0.80 & -0.77 & -0.55 & -0.41 & 2.09 \\ 0.57 & -0.17 & 0.28 & -0.66 & 0.85 & -0.68 & -1.30 & -0.39 & -2.27 & 1.37 & 0.07 & 2.29 \\ -0.48 & -2.28 & -0.93 & 1.34 & -0.26 & 0.35 & -0.14 & 0.17 & 0.46 & 1.54 & 0.85 & -1.22 \end{bmatrix}$$

$$\boldsymbol{\Gamma}_{Arn.} = \begin{bmatrix} 1.22 & 0.71 & 0.00 & -0.71 & -1.22 & -1.41 & -1.22 & -0.71 & 0.00 & 0.71 & 1.22 & 1.41 \\ 0.71 & 1.22 & 1.41 & 1.22 & 0.71 & 0.00 & -0.71 & -1.22 & -1.41 & -1.22 & -0.71 & 0.00 \\ 0.00 & -1.41 & 0.00 & 1.41 & 0.00 & -1.41 & 0.00 & 1.41 & 0.00 & -1.41 & 0.00 & 1.41 \\ 1.41 & 0.00 & -1.41 & 0.00 & 1.41 & 0.00 & -1.41 & 0.00 & 1.41 & 0.00 & -1.41 & 0.00 \\ -1.22 & 0.71 & 0.00 & -0.71 & 1.22 & -1.41 & 1.22 & -0.71 & 0.00 & 0.71 & -1.22 & 1.41 \\ 0.71 & -1.22 & 1.41 & -1.22 & 0.71 & 0.00 & -0.71 & 1.22 & -1.41 & 1.22 & -0.71 & 0.00 \end{bmatrix}$$

$$\boldsymbol{\Gamma}_{Art.} = \begin{bmatrix} 2.45 & -2.45 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2.45 & -2.45 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2.45 & -2.45 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2.45 & -2.45 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2.45 & -2.45 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2.45 & -2.45 \end{bmatrix}$$

Abbreviations: Latin Hypercube Sampling (LHS), Arndt (Arn.), Artavia et al. (Art.)

Source: own calculations

Box 3.10: Example LHS, Arndt, and Artavia et al. transformed quadratures (obtained via Cholesky decomposition) for the multivariate normal PDF inferred from the matrix of deviates

$$\mathbf{X}_{LHS} = \mathbf{L}\boldsymbol{\Gamma}_{LHS} = \begin{bmatrix} -0.024 & -0.054 & -0.087 & -0.064 & 0.053 & 0.076 & 0.030 & 0.011 & 0.128 & -0.123 & 0.078 & 0.000 \\ -0.035 & 0.036 & -0.181 & 0.174 & -0.066 & 0.113 & -0.216 & 0.125 & -0.022 & 0.430 & -0.095 & 0.002 \\ 0.048 & 0.002 & -0.070 & -0.025 & -0.009 & -0.042 & 0.072 & 0.006 & 0.045 & -0.080 & 0.069 & -0.015 \\ -0.124 & 0.026 & 0.012 & 0.024 & -0.004 & -0.130 & 0.091 & 0.050 & -0.089 & -0.011 & -0.024 & 0.119 \\ 0.086 & -0.026 & -0.082 & -0.029 & 0.053 & 0.051 & -0.163 & -0.006 & -0.077 & 0.170 & 0.027 & 0.104 \\ -0.027 & -0.064 & -0.046 & 0.002 & 0.016 & -0.013 & 0.027 & 0.010 & 0.014 & -0.012 & 0.050 & 0.015 \end{bmatrix}$$

$$\mathbf{X}_{Arn.} = \mathbf{L}\boldsymbol{\Gamma}_{Arn.} = \begin{bmatrix} 0.099 & 0.057 & 0.000 & -0.057 & -0.099 & -0.114 & -0.099 & -0.057 & 0.000 & 0.057 & 0.099 & 0.114 \\ 0.131 & 0.195 & 0.207 & 0.163 & 0.076 & -0.032 & -0.131 & -0.195 & -0.207 & -0.163 & -0.076 & 0.032 \\ 0.057 & -0.023 & -0.001 & 0.021 & -0.058 & -0.121 & -0.057 & 0.023 & 0.001 & -0.021 & 0.058 & 0.121 \\ 0.073 & -0.043 & -0.084 & 0.049 & 0.105 & -0.018 & -0.073 & 0.043 & 0.084 & -0.049 & -0.105 & 0.018 \\ -0.033 & 0.134 & 0.096 & -0.031 & 0.020 & -0.162 & 0.033 & -0.134 & -0.096 & 0.031 & -0.020 & 0.162 \\ 0.050 & -0.023 & 0.010 & -0.042 & 0.010 & -0.067 & -0.050 & 0.023 & -0.010 & 0.042 & -0.010 & 0.067 \end{bmatrix}$$

$$\mathbf{X}_{Art.} = \mathbf{L}\boldsymbol{\Gamma}_{Art.} = \begin{bmatrix} 0.198 & -0.198 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.055 & -0.055 & 0.358 & -0.358 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.114 & -0.114 & -0.002 & 0.002 & 0.096 & -0.096 & 0 & 0 & 0 & 0 & 0 & 0 \\ -0.032 & 0.032 & 0.006 & -0.006 & 0.063 & -0.063 & 0.151 & -0.151 & 0 & 0 & 0 & 0 \\ 0.115 & -0.115 & 0.103 & -0.103 & -0.001 & 0.001 & -0.063 & 0.063 & 0.167 & -0.167 & 0 & 0 \\ 0.064 & -0.064 & -0.009 & 0.009 & 0.027 & -0.027 & 0.029 & -0.029 & 0.024 & -0.024 & 0.056 & -0.056 \end{bmatrix}$$

Abbreviations: Latin Hypercube Sampling (LHS), Arndt (Arn.), Artavia et al. (Art.)

Source: own calculations

Box 3.11: Example: LHS, Arndt, and Artavia et al. transformed quadratures (obtained via the diagonalization method) for the multivariate normal PDF inferred from the matrix of deviates

$$\mathbf{X}_{LHS} = \mathbf{U}\sqrt{\mathbf{D}}\boldsymbol{\Gamma}_{LHS} = \begin{bmatrix} 0.105 & -0.018 & 0.032 & -0.125 & 0.042 & -0.037 & 0.093 & -0.077 & 0.045 & -0.237 & 0.100 & -0.016 \\ -0.004 & -0.072 & -0.209 & -0.046 & 0.049 & 0.149 & -0.015 & 0.048 & 0.206 & -0.061 & 0.111 & -0.031 \\ 0.054 & 0.005 & 0.005 & -0.050 & -0.016 & -0.053 & 0.138 & -0.030 & 0.080 & -0.203 & 0.070 & -0.076 \\ 0.016 & 0.071 & -0.028 & 0.039 & -0.052 & -0.156 & 0.069 & 0.035 & -0.102 & 0.046 & -0.006 & 0.060 \\ -0.161 & -0.086 & -0.001 & -0.127 & 0.121 & 0.068 & 0.091 & 0.001 & 0.124 & -0.268 & 0.068 & 0.091 \\ 0.002 & -0.043 & -0.005 & -0.012 & 0.004 & -0.057 & 0.071 & -0.012 & 0.000 & -0.070 & 0.053 & -0.002 \end{bmatrix}$$

$$\mathbf{X}_{Arn.} = \mathbf{U}\sqrt{\mathbf{D}}\boldsymbol{\Gamma}_{Arn.} = \begin{bmatrix} -0.064 & -0.066 & -0.069 & -0.084 & -0.104 & -0.089 & 0.064 & 0.066 & 0.069 & 0.084 & 0.104 & 0.089 \\ 0.189 & 0.138 & 0.083 & -0.030 & -0.155 & -0.212 & -0.189 & -0.138 & -0.083 & 0.030 & 0.155 & 0.212 \\ 0.010 & -0.097 & -0.056 & -0.007 & -0.091 & -0.035 & -0.010 & 0.097 & 0.056 & 0.007 & 0.091 & 0.035 \\ 0.033 & -0.052 & -0.023 & 0.113 & 0.067 & -0.078 & -0.033 & 0.052 & 0.023 & -0.113 & -0.067 & 0.078 \\ 0.106 & 0.044 & -0.118 & -0.144 & -0.069 & -0.057 & -0.106 & -0.044 & 0.118 & 0.144 & 0.069 & 0.057 \\ 0.011 & -0.074 & -0.024 & -0.037 & 0.003 & -0.044 & -0.011 & 0.074 & 0.024 & 0.037 & -0.003 & 0.044 \end{bmatrix}$$

$$\mathbf{X}_{Art.}\mathbf{U}\sqrt{\mathbf{D}}\boldsymbol{\Gamma}_{Art.} = \begin{bmatrix} 0.077 & -0.077 & -0.163 & 0.163 & 0.041 & -0.041 & -0.057 & 0.057 & 0.037 & -0.037 & -0.013 & 0.013 \\ 0.343 & -0.343 & 0.112 & -0.112 & 0.026 & -0.026 & -0.028 & 0.028 & -0.001 & 0.001 & 0.004 & -0.004 \\ 0.044 & -0.044 & -0.108 & 0.108 & 0.072 & -0.072 & -0.015 & 0.015 & -0.056 & 0.056 & -0.004 & 0.004 \\ -0.020 & 0.020 & 0.046 & -0.046 & 0.140 & -0.140 & 0.071 & -0.071 & 0.015 & -0.015 & -0.015 & 0.015 \\ 0.174 & -0.174 & -0.108 & 0.108 & -0.076 & 0.076 & 0.090 & -0.090 & 0.000 & 0.000 & -0.007 & 0.007 \\ 0.019 & -0.019 & -0.065 & 0.065 & 0.047 & -0.047 & 0.022 & -0.022 & 0.011 & -0.011 & 0.047 & -0.047 \end{bmatrix}$$

Abbreviations: Latin Hypercube Sampling (LHS), Arndt (Arn.), Artavia et al. (Art.)

Source: own calculations

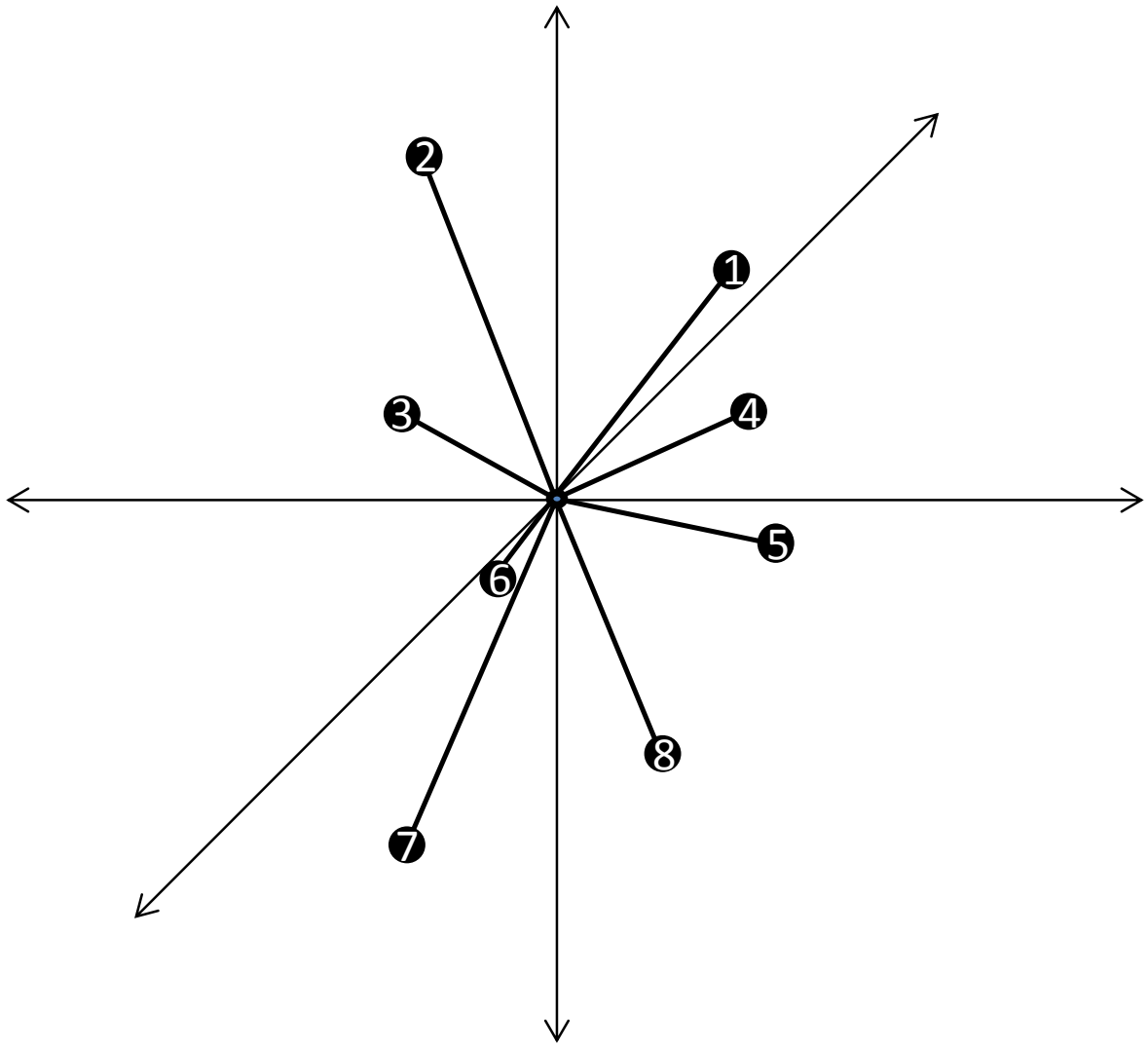
Boxes 3.7 to 3.11 show that the procedure of the generation of degree 3 quadratures can be divided into 4 steps:

1. computation of deviates, $\mathbf{Z}$, and of the covariance matrix, $\boldsymbol{\Sigma}[\boldsymbol{z}]$;
2. factorization of the covariance matrix in the form $\boldsymbol{\Sigma}[\boldsymbol{z}] = \mathbf{A}\mathbf{A}^{\mathrm{T}}$;
3. generation of standard normal quadratures, $\boldsymbol{\Gamma}$, via LHS, Arndt's formula or Artavia et al.'s formula;
4. production of the final quadratures with the desired covariance by computing $\mathbf{A}\boldsymbol{\Gamma}$.

One important difference is brought into light in Box 3.9. It can be observed how the PDF of the stochastic variables is approximated with a major diversity of values with the LHS method than with Arndt's approach and with Artavia et al.'s formula. This is obtained as a result of the allocation of the quadrature points in space. With LHS, the quadratures are distributed all around the domain and with Arndt's and Artavia et al.'s formulae these are fully symmetrical. Figures 3.2, 3.3 and 3.4 give a graphical interpretation of the quadrature points obtained with the different approaches for $n = 3$. The symmetry of the Q^n is shown in Figures 3.2 and 3.3. Figure 3.4 below exposes an interpretation of the distribution of points obtained with LHS. In it, the 3-dimensional space has been divided in 8 quadrants of equal volume (equal probability) and one point has been randomly assigned to each quadrant. Note that the diversity of values obtained with Artavia et al. is the lowest, since the vertices of the Q^n lie on the coordinate axes. In this way, the PDF of the stochastic variables is approximated only with two values different from zero, namely $\pm\sqrt{6}$.

The phenomenon of little variability of values in the marginal approximations when using Artavia et al. is partially stopped with the induction of the covariance matrix. That step implies a rotation and extension of the standard normal quadratures and with it, the vertices do not lie on the coordinate axes anymore. With the rotation obtained with the diagonalization method, all coordinates will have a value different from zero. With the Cholesky decomposition, the discrete approximation still presents some zeros (see Box 3.10). This occurs because the lower triangular matrix, **L**, also presents zeros which are carried on with the multiplication of the matrices.

Figure 3.4: Interpretation of the distribution of points obtained with Latin Hypercube Sampling (LHS)



Source: own development

Box 3.12: Example accuracy of approximation of descriptive statistics of the quadratures generated via Cholesky decomposition

Descriptive statistics of the historical data and the simulated values of the 6 stochastic variables:

Historical Data

Mean	*SD*	*Min*	*Max*
0	0.081	−0.130	0.116
0	0.148	−0.348	0.187
0	0.061	−0.148	0.095
0	0.068	−0.122	0.116
0	0.096	−0.161	0.177
0	0.040	−0.071	0.069

LHS:				*Arn.*:				*Art.*:			
Mean	*SD*	*Min*	*Max*	*Mean*	*SD*	*Min*	*Max*	*Mean*	*SD*	*Min*	*Max*
0.002	0.072	−0.123	0.128	0	0.081	−0.114	0.114	0	0.081	−0.198	0.198
0.022	0.167	−0.216	0.430	0	0.148	−0.207	0.207	0	0.148	−0.358	0.358
0.000	0.049	−0.080	0.072	0	0.061	−0.121	0.121	0	0.061	−0.114	0.114
−0.005	0.075	−0.130	0.119	0	0.068	−0.105	0.105	0	0.068	−0.151	0.151
0.009	0.088	−0.163	0.170	0	0.096	−0.162	0.162	0	0.096	−0.167	0.167
−0.002	0.030	−0.064	0.050	0	0.040	−0.067	0.067	0	0.040	−0.064	0.064

Differences with respect to historical data:

LHS:				*Arn.*:				*Art.*:			
Mean	*SD*	*Min*	*Max*	*Mean*	*SD*	*Min*	*Max*	*Mean*	*SD*	*Min*	*Max*
0.002	−0.009	0.007	0.012	0	0	0.016	−0.002	0	0	−0.068	0.082
0.022	0.019	0.132	0.243	0	0	0.141	0.019	0	0	−0.010	0.171
0.000	−0.012	0.068	−0.023	0	0	0.027	0.027	0	0	0.034	0.020
−0.005	0.007	−0.009	0.003	0	0	0.017	−0.011	0	0	−0.029	0.035
0.009	−0.008	−0.002	−0.007	0	0	−0.001	−0.015	0	0	−0.006	−0.010
−0.002	−0.009	0.008	−0.019	0	0	0.005	−0.003	0	0	0.007	−0.005

Abbreviations: Latin Hypercube Sampling (LHS), Arndt (Arn.), Artavia et al. (Art.), Standard deviation (SD)

Source: own calculations

Box 3.13: Example accuracy of approximation of the correlation of the quadratures generated via Cholesky decomposition

$$\mathbf{P}[\mathbf{Z}] = \begin{bmatrix} 1 & 0.152 & 0.767 & -0.192 & 0.487 & 0.660 \\ 0.152 & 1 & 0.102 & 0.004 & 0.505 & 0.005 \\ 0.767 & 0.102 & 1 & 0.097 & 0.363 & 0.689 \\ -0.192 & 0.004 & 0.097 & 1 & -0.321 & 0.245 \\ 0.487 & 0.505 & 0.363 & -0.321 & 1 & 0.373 \\ 0.660 & 0.005 & 0.689 & 0.245 & 0.373 & 1 \end{bmatrix}$$

$$\mathbf{P}[\mathbf{X}_{LHS}] = \begin{bmatrix} 1 & -0.419 & 0.621 & -0.304 & -0.248 & 0.605 \\ -0.419 & 1 & -0.559 & -0.131 & 0.640 & -0.157 \\ 0.621 & -0.559 & 1 & -0.038 & -0.401 & 0.506 \\ -0.304 & -0.131 & -0.038 & 1 & -0.230 & 0.185 \\ -0.248 & 0.640 & -0.401 & -0.230 & 1 & -0.041 \\ 0.605 & -0.157 & 0.506 & 0.185 & -0.041 & 1 \end{bmatrix}$$

$$\mathbf{P}[\mathbf{X}_{LHS}] - \mathbf{P}[\mathbf{Z}] = \begin{bmatrix} 0 & -0.571 & -0.145 & -0.113 & -0.735 & -0.055 \\ -0.571 & 0 & -0.661 & -0.135 & 0.136 & -0.162 \\ -0.145 & -0.661 & 0 & -0.135 & -0.764 & -0.183 \\ -0.113 & -0.135 & -0.135 & 0 & 0.092 & -0.060 \\ -0.735 & 0.136 & -0.764 & 0.092 & 0 & -0.413 \\ -0.055 & -0.162 & -0.183 & -0.060 & -0.413 & 0 \end{bmatrix}$$

$$\mathbf{P}[\mathbf{X}_{Arn.}] = \begin{bmatrix} 1 & 0.152 & 0.767 & -0.192 & 0.487 & 0.660 \\ 0.152 & 1 & 0.102 & 0.004 & 0.505 & 0.005 \\ 0.767 & 0.102 & 1 & 0.097 & 0.363 & 0.689 \\ -0.192 & 0.004 & 0.097 & 1 & -0.321 & 0.245 \\ 0.487 & 0.505 & 0.363 & -0.321 & 1 & 0.373 \\ 0.660 & 0.005 & 0.689 & 0.245 & 0.373 & 1 \end{bmatrix}$$

$$\mathbf{P}[\mathbf{X}_{Arn.}] - \mathbf{P}[\mathbf{Z}] = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\mathbf{P}[\mathbf{X}_{Art.}] = \begin{bmatrix} 1 & 0.152 & 0.767 & -0.192 & 0.487 & 0.660 \\ 0.152 & 1 & 0.102 & 0.004 & 0.505 & 0.005 \\ 0.767 & 0.102 & 1 & 0.097 & 0.363 & 0.689 \\ -0.192 & 0.004 & 0.097 & 1 & -0.321 & 0.245 \\ 0.487 & 0.505 & 0.363 & -0.321 & 1 & 0.373 \\ 0.660 & 0.005 & 0.689 & 0.245 & 0.373 & 1 \end{bmatrix}$$

$$\mathbf{P}[\mathbf{X}_{Art.}] - \mathbf{P}[\mathbf{Z}] = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Abbreviations: Latin Hypercube Sampling (LHS), Arndt (Arn.), Artavia et al. (Art.), Standard deviation (SD)

Source: own calculations

Box 3.14: Example inferred PDF of the deviates and simulated frequency distribution of the LHS, Arndt, and Artavia et al. quadratures: the case of the second variable

Inferred PDF of the second variable of the matrix of deviates (in %): $x_2 \sim N(0, 225)$ (class limits are $\pm\sigma_2$):

Class No.	*Class limits*	*Probability*
1	< -15	0.159
2	-15 *to* 15	0.683
3	> 15	0.159

Frequency distribution of the second variable of the transformed quadratures (the rounded distance from the minimum to the maximum value is divided into 4 regions of equal distance, class 1 = the 1st region, class 2 = the 2nd and 3rd region, class 3 = 4th region):

CHOLESKY:

LHS:			*Arn.*:			*Art.*:		
Class No.	*Class limits*	*Probability*	*Class No.*	*Class limits*	*Probability*	*Class No.*	*Class limits*	*Probability*
1	-37.5 *to* -12.5	0.167	1	-22.5 *to* -7.5	0.417	1	-37.5 *to* -12.5	0.083
2	-12.5 *to* 12.5	0.583	2	-7.5 *to* 7.5	0.167	2	-12.5 *to* 12.5	0.833
3	12.5 *to* 37.5	0.250	3	7.5 *to* 22.5	0.417	3	12.5 *to* 37.5	0.083

DIAGONAL.:

LHS:			*Arn.*:			*Art.*:		
Class No.	*Class limits*	*Probability*	*Class No.*	*Class limits*	*Probability*	*Class No.*	*Class limits*	*Probability*
1	-22.5 *to* -7.5	0.083	1	-22.5 *to* -7.5	0.417	1	-37.5 *to* -12.5	0.083
2	-7.5 *to* 7.5	0.667	2	-7.5 *to* 7.5	0.167	2	-12.5 *to* 12.5	0.833
3	7.5 *to* 22.5	0.250	3	7.5 *to* 22.5	0.417	3	12.5 *to* 37.5	0.083

Abbreviations: Latin Hypercube Sampling (LHS), Arndt (Arn.), Artavia et al. (Art.), Standard deviation (SD)

Source: own calculations

Boxes 3.12 to 3.14 analyze the quality of the degree of precision of the quadratures for $\mathbf{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with respect to the following parameters:

1. precision of approximation of the mean and variance;
2. precision of approximation of the correlation between the stochastic variables;
3. frequency distribution of the generated points.

With respect to the first parameter, in Box 3.12 it can be observed that Arndt‘s and Artavia et al.’s quadratures approximate accurately the desired means and the variances since there is no difference with those of the historical data. The quadratures obtained with LHS match the means well, but the approximations of the standard deviation are regular. This occurs due to the random characteristic of this method and the small number of points used for the 6-dimensional integration problem. Even though it partitions the multivariate normal PDF into 12 sub-regions of equal probability, this partitioning is not enough to approximate the first and second moments more accurately. Note that, compared to 'efficient quadratures', LHS requires much more points to obtain a similar degree of precision. This can also be seen in Figure 3.4, where one point must approximate $1/8$ of the 3-D region. For the 6-dimensional case of this example, the number $N = 12$ of generated points is too small for the random condition. Nevertheless, the size of the sample was chosen to be 12 for the comparison purposes of the exercise done in this section. The quadratures obtained with the other two methods are so accurate due to their systematic selection of points taking advantage of the symmetry of $\mathbf{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. In terms of accuracy of approximation of the mean and the variance, the quadratures obtained via Cholesky decomposition and via the Diagonalization method were practically equal; thus, only the results developed via Cholesky are shown.

In relation to the precision of the approximation of the correlation, it can be observed in Box 3.13 that again the LHS method is not precise, while Arndt and Artavia et al. match accurately the desired property. In this case, the random property of LHS makes it impossible to reproduce perfectly the correlation. Nonetheless, when computing t-student values (99% of confidence) it has failed to reject the null hypothesis that the simulated multivariate distribution statistically has the same correlation as the historical series, which indicates a similar correlation of the quadratures to the desired one. With Arndt and Artavia et al., the quadratures in $\boldsymbol{\Gamma}$ are orthogonal (perpendicular to each other); then, the rotation information in the factor matrix $\mathbf{A}$, is perfectly given to the matrix of final quadratures. With LHS, the vectors in $\boldsymbol{\Gamma}$ are not perfectly perpendicular to each other and thus, after transformation, the obtained correlation is not exactly equal to the desired one.

Finally, concerning the simulated frequency distribution, in Box 3.14 it can be observed that the LHS approximates more accurately the inferred distribution from the deviates than Arndt and Artavia et al. The frequency distribution of classes obtained with the LHS, as well as their probability, is very close to the desired one. For example, note that the width of class 2 is almost $[-\sigma, \sigma]$ as desired. Conversely, the quadratures obtained with the Arndt formula present problems to get the widths of the classes right – e.g. the central class 2 is very thin – and the majority of the points are allocated in classes 1 and 3 with values close to the desired standard deviation of 0.148. The Artavia et al. formula reproduces the class widths better than Arndt because it generates two values (± 0.358 in the quadratures with Cholesky and ± 0.343 in the quadratures with the diagonalization method) which are distant from the mean. These two values obtained with each transformation – Cholesky and diagonalization – are further away from the mean than two times the desired standard deviation (2σ). As a result, the rest of the values must be close to the mean in order to be able to approximate the first and second moment of the PDF of that variable precisely. Thus, class 2 has a higher probability than desired with a majority of points close to the value of zero (see boxes 3.10 and 3.11).

Moreover, note that the frequency distribution obtained via Cholesky and via the diagonalization method are equal. However, this is not the case for all variables. For example, in the case of the approximation of the first variable, with $\mathbf{L}\boldsymbol{\Gamma}_{Art.}$ and with $\mathbf{U}\sqrt{\mathbf{D}}\boldsymbol{\Gamma}_{Art.}$ different probability distributions are obtained. With $\mathbf{L}\boldsymbol{\Gamma}_{Art.}$ two extreme values, ± 0.198, and zeros are generated. With $\mathbf{U}\sqrt{\mathbf{D}}\boldsymbol{\Gamma}_{Art.}$ the zeros disappear and the points are better distributed along the probability classes.

This last analysis points out how the purpose of Arndt's and Artavia et al.'s quadratures is to approximate the first and second moments of the multivariate PDF precisely and not to distribute the points in accordance to a normal distribution. The LHS is less efficient and less precise in the simulation of the first and second moments; nonetheless, the frequency distribution of the points reproduces better the inferred PDF.

4 Documentation of the Stochastic Version of the European Simulation Model (ESIM)

The analysis of the consequences of different quadrature formulas on the approximation of the moments of model results is done with the stochastic version of ESIM. This version has been continuously developed; however, not all development steps have been described. This chapter gives a complete documentation of the works done for the stochastic version of the model up to December 2011.

4.1 Brief Description of ESIM

ESIM is a comparative static, partial equilibrium, net-trade, world market model of production and consumption of agricultural products, and of some first-stage processing activities (Banse et al., 2005). ESIM is a partial model, as only a part of the economy – the agricultural sector – is modeled, e.g., macroeconomic variables (such as income or real exchange rates) are exogenous. As a world market model, it includes all countries, though in greatly varying degrees of disaggregation. In the version used for this study, the Member States of the European Union (EU), Turkey, and the US are modeled as individual countries; all others countries of the world are aggregated in one region called "rest of the world" (ROW). ESIM has rich cross-commodity relations, as well as a detailed representation of EU policies; it depicts price and trade policy instruments, as well as direct payments. As ESIM is mainly designed to simulate the development of agricultural markets in the EU and accession candidates, policies are only modeled for these countries. For the US and the ROW, production and consumption take place at world market prices. Area allocation, yield and demand functions are isoelastic[6].

For an overview on the product coverage, the structure, and the equations in ESIM, please see Annexes 3.1 to 3.3.

[6] This first paragraph has been taken from Artavia et al. (2009). Some small changes have been made.

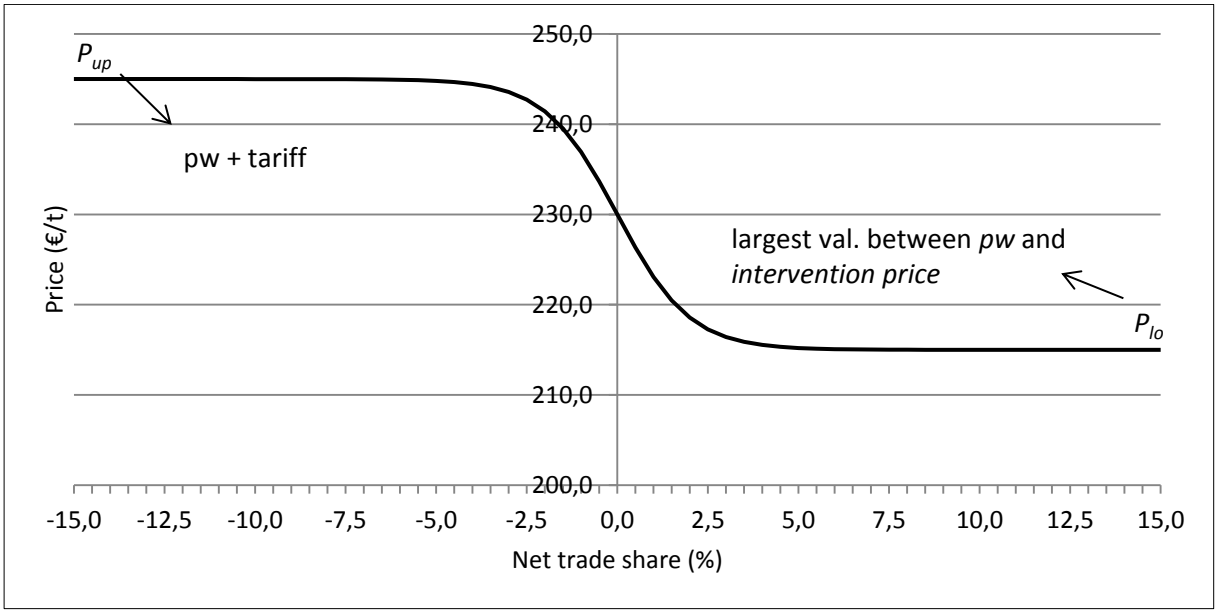
In addition to the basic information from above, it is important to understand the price system in ESIM for the analysis of results.

In ESIM, trade policies (tariffs, TRQs, export subsidies and export subsidy limits) are incorporated into the EU price equation. In order to simulate smooth transitions at threshold points between polices, ESIM uses a logistical functional form. For example, the transition from TRQs to full import tariffs is smooth and controlled by the parameters of the logistic function.

In order to capture all possible policies, ESIM applies a system of two logistic functions. The upper bounds of those functions depict tariff or export subsidy price levels. The lower bound P_{lo} is the same for both functions and is determined by the largest value between the intervention price – if it applies – and the world market price.

In the logistic function considering tariffs, the upper price level P_{up} is determined by the largest value between the world market price plus tariff (or threshold price) and the world market price alone. This is exemplified in Figure 4.1. In the case that no tariffs are applied, then the logistic function will be a horizontal line at the P_{lo} level.

Figure 4.1: Example of the price logistic function in ESIM considering tariffs and intervention prices



Note: (P_{up}) denotes the upper prices level of the logistic function, (P_{lo}) denotes the lower price level of the logistic function.

Source: own development based on Banse et al. (2005)

Figure 4.1 shows how in the range of $[-5{,}5]$ percent of net trade share (with respect to total demand), the price is between the upper and lower price bound as desired. This logistic function is defined as follows (Banse et al. 2005):

$$PD = \left(P_{up} - P_{lo}\right) \cdot \frac{-\alpha \cdot e^{\beta \cdot TRADSHR}}{1 + \alpha \cdot e^{\beta \cdot TRADSHR}} + P_{up} \tag{4.1}$$

where,

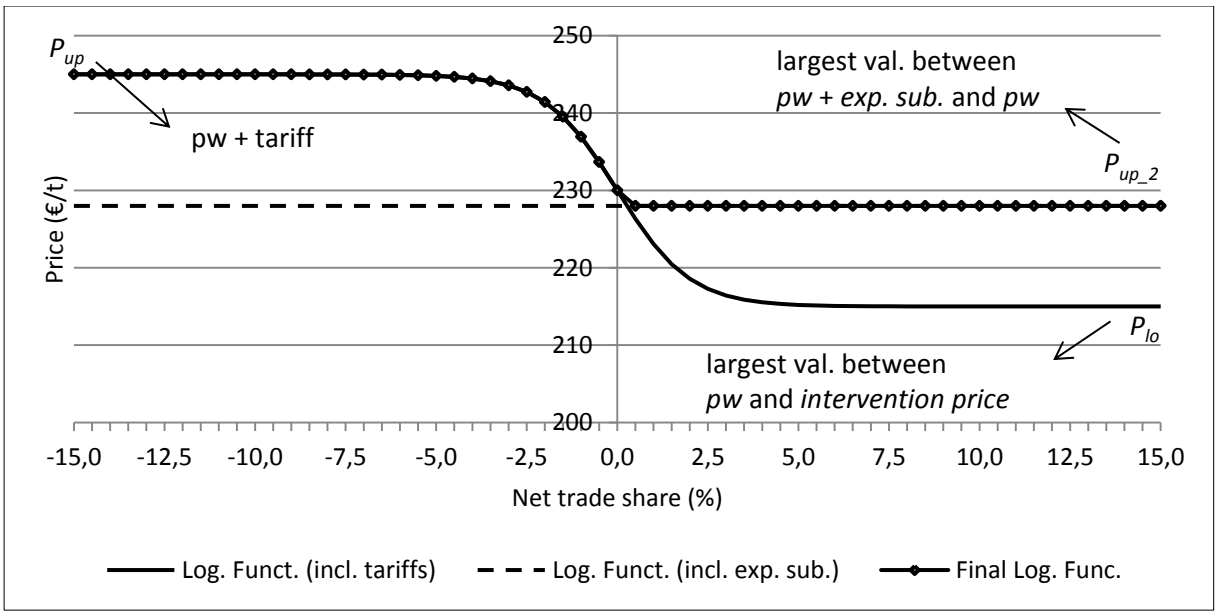
$$TRADSHR = \frac{NX}{TUSE} \cdot 100 \tag{4.2}$$

and α and $\beta = 1$.

In the logistic function considering export subsidies, the upper price level is determined by the largest value between the world market price plus the subsidy and the world market price alone.

Figure 4.2 shows the price system for products subject to tariffs and export subsidies. In those cases, both logistic functions in ESIM are required and the largest value between them is taken as the final price curve. To be able to distinguish between the two upper price bounds, the level determined by the export subsidy is named P_{up_2}.

Figure 4.2: Example of the price mechanism in ESIM for products including tariffs and export subsidies

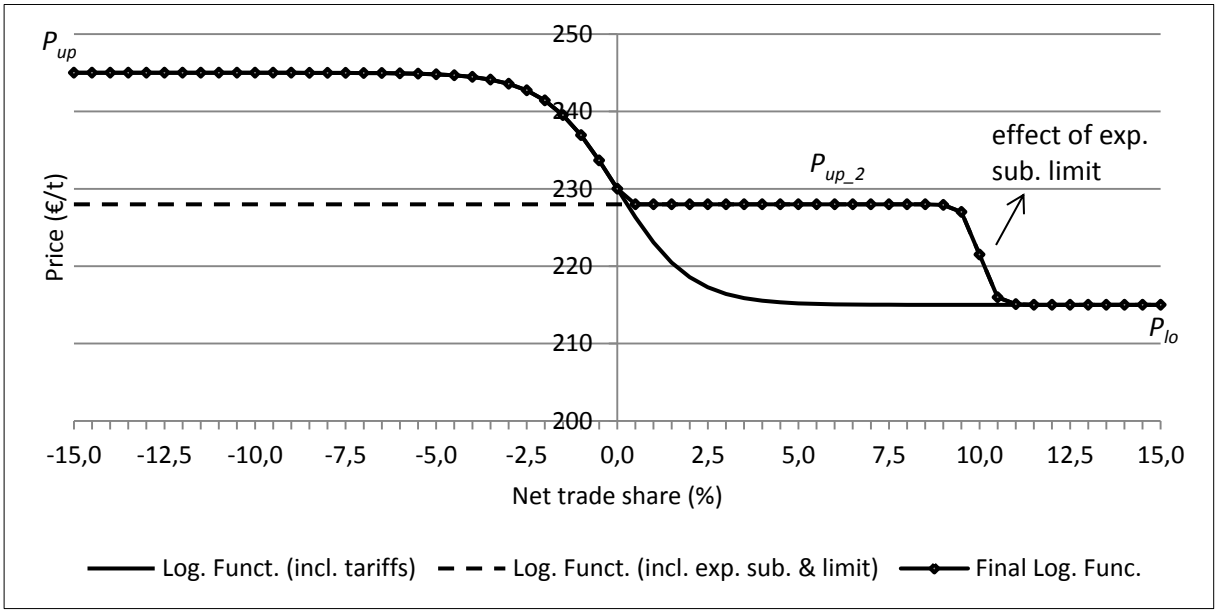


Note: (P_{up}) denotes the upper prices level of the logistic function including tariffs, (Pup_2) denotes the upper prices level of the logistic function including export subsidies, (P_{lo}) denotes the lower price level of both logistic functions, (exp.sub) denotes export subsidy.

Source: own development based on Banse et al. (2005)

In Figure 4.2, it can be observed how: 1) in the net import situation, the final price takes the highest price level; 2) in the range close to zero net trade, the price decreases smoothly; and 3) in the export situation, the final price takes the price level considering the export subsidy. Note that the logistic function including export subsidies is a horizontal line above P_{lo}, indicating that there is no quantity limit on the subsidies. However, due to WTO restrictions, export subsidies are normally subjected to quantity limits. Moreover, for certain products, it is common in the EU to grants TRQs – import quotas with lower tariffs levels or duty free. These two policies can also be integrated in the logistic functions. Figures 4.3 and 4.4 show how such policies affect the price building system.

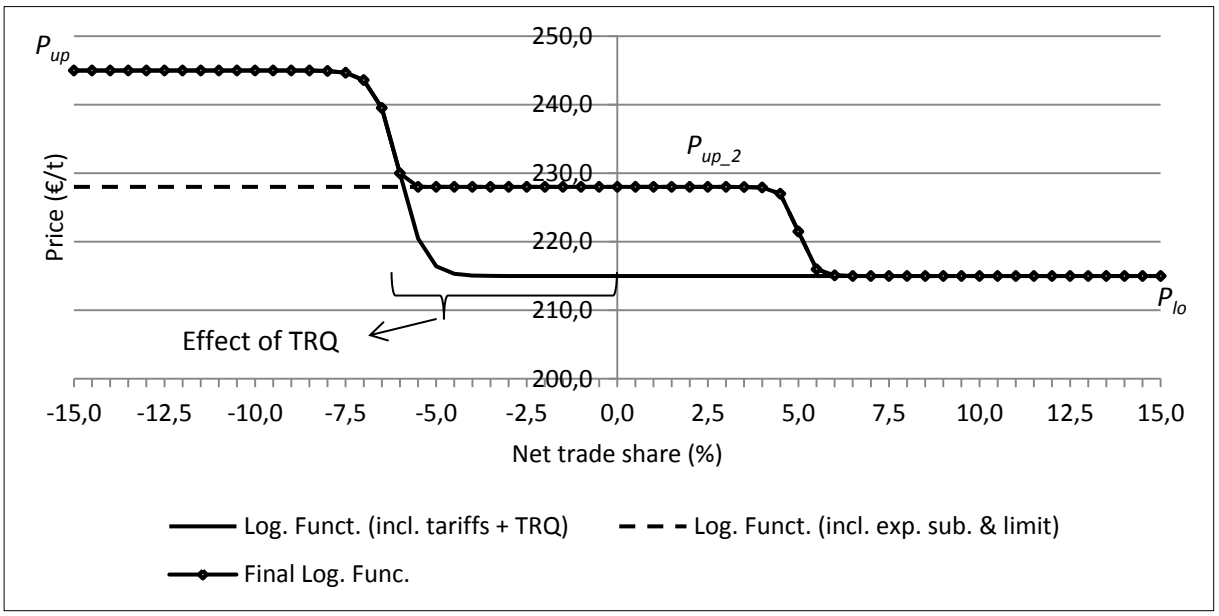
Figure 4.3: Example of the price mechanism in ESIM for products including tariffs, export subsidies, and export subsidy limits



Note: (P_{up}) denotes the upper prices level of the logistic function including tariffs, (P_{up_2}) denotes the upper prices level of the logistic function including export subsidies, (P_{lo}) denotes the lower price level of both logistic functions, (exp.sub) denotes export subsidy.

Source: own development based on Banse et al. (2005)

Figure 4.4: Example of the price mechanism in ESIM for products including tariffs, export subsidies, export subsidy limits, and TRQs



Note: (P_{up}) denotes the upper prices level of the logistic function including tariffs, (P_{up_2}) denotes the upper prices level of the logistic function including export subsidies, (P_{lo}) denotes the lower price level of both logistic functions, (exp.sub) denotes export subsidy.

Source: own development based on Banse et al. (2005)

Figure 4.3 shows the effect of export subsidy limits and Figure 4.4 shows the effect of both, export subsidy limits and TRQs. In Figure 4.3, it can be seen how the export subsidy limit results in a smooth decrease of the EU domestic price to world market levels. In Figure 4.4, it can be observed how for commodities with both policies, the P_{up_2} applies until the TRQ is exhausted (net trade share of - 6% in the example).The logistic function including export subsidies and export subsidy limits is given by (Banse et al., 2005):

$$PD = \left(P_{up\_2} - P_{lo}\right) \cdot \frac{-\alpha \cdot e^{\beta \cdot TRADSHR}}{1 + \alpha \cdot e^{\beta \cdot TRADSHR}} + P_{up\_2} \tag{4.3}$$

where,

$$TRADSHR = \frac{NX - subsquant}{TUSE} \cdot 100 \tag{4.4}$$

and $subsquant$ is the export subsidy limit (in quantities).

The logistic function for the curve including tariffs and TRQ is given by (own development):

$$PD = \left(P_{up} - P_{lo}\right) \cdot \frac{-\alpha \cdot e^{\beta \cdot TRADSHR}}{1 + \alpha \cdot e^{\beta \cdot TRADSHR}} + P_{up} \tag{4.5}$$

where,

$$TRADSHR = \frac{NX + TRQ}{TUSE} \cdot 100. \tag{4.6}$$

Note that the slope of the smooth transition between price levels is controlled by the parameter beta β in the functions. This enables the simulation of slower or faster transitions as desired.

4.2 Considered Uncertainty and Basic Adaptations

In the stochastic version of ESIM up to December 2011, the yields of wheat, barley and rapeseed in all countries/regions are considered as stochastic, accounting for their uncertainty around linear trends. The stochastic terms simulating that uncertainty are incorporated into the respective supply and yield equations. The crops were selected due to their importance in the EU. All three crops are significant in terms of their share of production in the EU, and wheat and rapeseed especially significant in terms of their political importance; wheat as a sensitive food product and rapeseed as the main feedstock for the production of biodiesel.

Due to its focus on the EU, in ESIM, the equations for crop supply for the EU Member States are divided in area and yield equations, supply then being obtained by:

$$SUPPLY = AREA \cdot YIELD. \tag{4.7}$$

Crops supply for the rest of countries and regions – US and the ROW – is only one isoelastic function:

$$SUPPLY = \text{sup. int} \cdot \prod_{crops} PP^{elastsp_{crops,crops}} \cdot \text{tp. gr.} \tag{4.8}$$

For the EU Member States, crop yields are modeled with isoelastic functions dependent on producer prices, on intermediate cost indices and on technical progress, as follows:

$$YIELD = \text{yild. int} \cdot PP^{elastyd} \cdot \text{int. ind}^{elastyi} \cdot \text{lab. ind}^{elastyl} \cdot \text{tp. gr.} \tag{4.9}$$

Note that the yield equations are only subject to own price elasticities. The cross price elasticities are considered in the area equations which are not presented here.

For the stochastic version, the stochastic variable $stoch$ has been included for the corresponding crops in all regions as a factor into (4.8) and (4.9), as follows.

For the US and the ROW:

$$SUPPLY = \text{sup.int} \cdot \prod_{crops} PP^{elastsp_{crops,crops}} \cdot \text{tp.gr} \cdot (1 + stoch), \tag{4.10}$$

for the European countries:

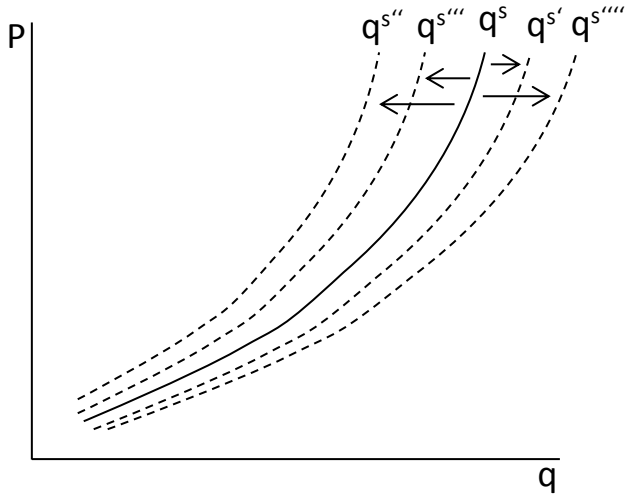
$$YIELD = \text{yild.int} \cdot PP^{elastyd} \cdot \text{int.ind}^{elastyi} \cdot \text{lab.ind}^{elastyl} \cdot \text{tp.gr} \cdot (1 + stoch). \tag{4.11}$$

$stoch$ contains the generated quadratures for $\mathbf{x} \sim N(\mathbf{0}, \mathbf{\Sigma})$, which is the assumed stochastic space of the random vector.

Note that as mentioned before, the stochastic terms capture the uncertainty of yields around linear trends. Thus, factors such as e.g. product and input prices or structural changes are not captured by the trend and these may explain some of the variability of yields. However, an econometric model with a higher number of independent variables for the crop/country disaggregation in ESIM presents high data requirements, which are outside the scope of this study. Still, the uncertainty around the trend is considered in this study as an approximation of the influence of weather on yields, since it is an important factor responsible for annual yield fluctuations.

By running the model several times over the quadrature points contained in $stoch$, the consequences of yield uncertainty on the stability of markets can be studied. Figure 4.5 shows the effect of the stochastic factor on the crop supply curves. It can be observed how the factor shifts the curve to the left and to the right.

Figure 4.5: Effect of different values of the stochastic variables on the crop supply curves



Source: own development

4.3 Calibration of Supply Elasticities

In its deterministic version, ESIM has been developed to simulate the medium run adaptation of agricultural markets to shocks to the system (e.g. different policy or structural change scenarios). This condition cannot be used for the stochastic version, since it is assumed that the shocks introduced by the stochastic terms capture mainly the weather effect on yields. Such shocks require the simulation of the short run adaptation capacity of agricultural markets, which is very limited.

To simulate that short run market reaction to price changes, the most appropriate would be a stock function. Anyhow, due to its comparative static quality and former deterministic condition, ESIM was not foreseen with such a function. Therefore, the short run reaction capacity of the agricultural markets is simulated through a re-calibration of the crop and livestock supply curves with reduced output and input price elasticities.

More specifically, the calibration of the supply functions for the stochastic runs is computed with the values of supply, price and cost indexes of the last simulation year of the baseline scenario – which is 2015 for the version used here – and with the reduced elasticities. For example, the calibration of the new yield curve for crops in the EU Member States is computed as follows:

$$\text{yild.int} = \frac{YIELD}{PP^{elastyd}\,\text{int.ind}^{elastyi}\,\text{lab.ind}^{elastyl}}. \tag{4.12}$$

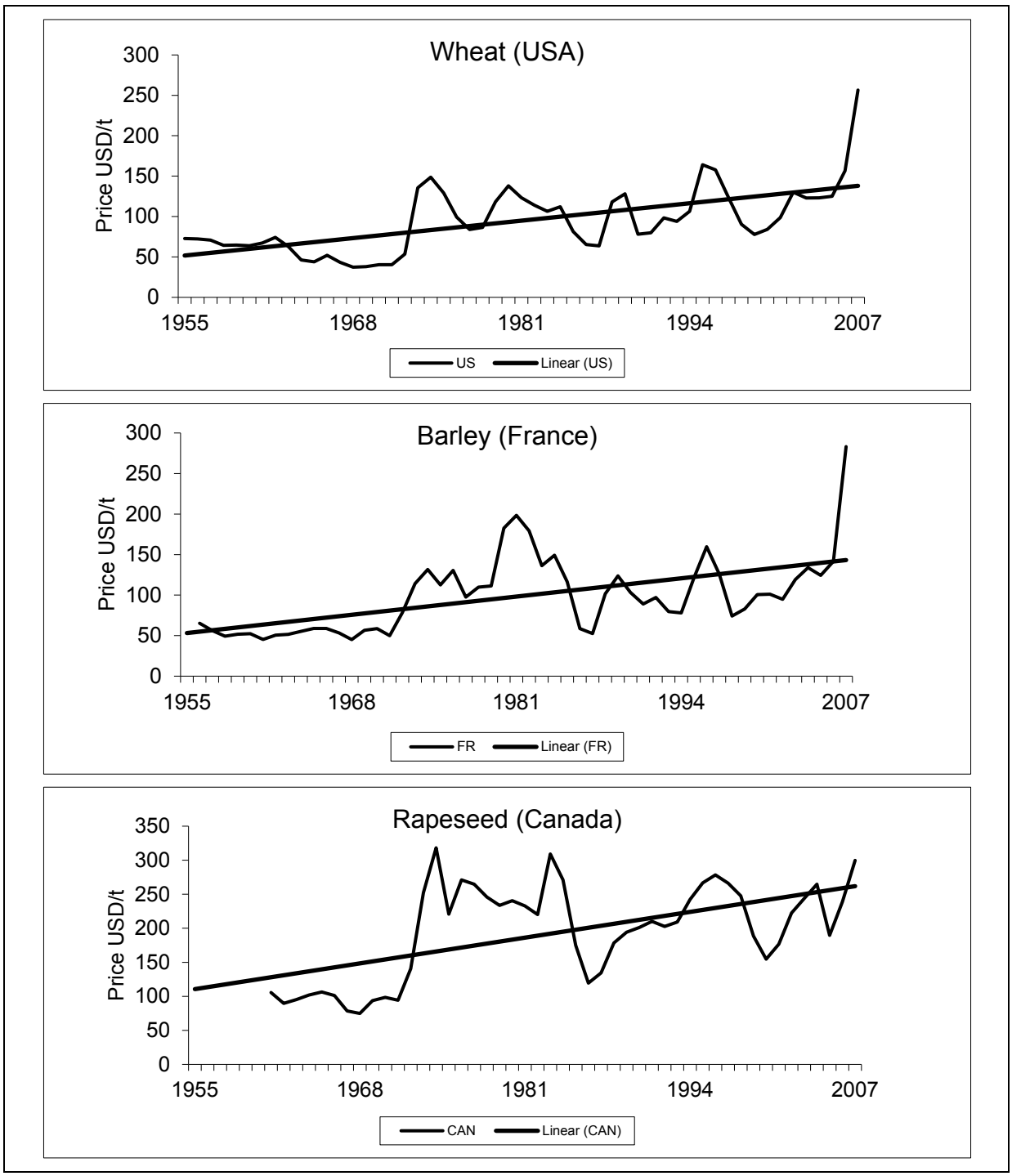
A detail which should not be forgotten is the shifter $tp.gr$ which is included in the supply equations; see (4.10) and (4.11). For the calibration, $tp.gr$ must be set to 1 in order to be able to reproduce the quantity and price values of 2015. If not, $tp.gr$ shifts the curve to the right and ESIM generates new equilibrium quantities and prices which differ from those used for the calibration with reduced elasticities. Besides, for the stochastic runs, technical progress must remain off – set equal to one – since no progress is expected on the short run (intra-annual).

The demand side is left with its original calibration for almost all products, since consumption is already modeled as short run behavior to price changes. Furthermore, it is expected that consumption reaction will not change much between the short and the medium run. Changes are made only for the products biodiesel and bioethanol, where, for the stochastic runs, their human demand is fixed to the deterministic level of 2015. This is done with the intention of correcting the simulation of a too elastic consumption behavior, since the production side – due to some required time for processing activities – cannot respond to those rapid demand changes. More specifically, it corrects for a too small rapeseed price reaction to the stochastic changes, which is obtained with highly elastic biodiesel demand.

Since there was no reference available on the value of short run supply elasticities, ESIM is calibrated in such a way that it reproduces nominal price uncertainty observed in the past. For that purpose, the database from Anderson and Valenzuela (2008) is used, which offers a time series (1955 to 2007) of boarder prices. For the analysis, it must be kept in mind that the simulated price instability with ESIM is an approximation of total price variability, which includes several sources of volatility, such as e.g., weather, crude oil price, financial markets, world demand and crop expansion. One attempt to overcome this problem in future research projects is to intend to capture the main sources of uncertainty in the modeling framework.

For the analysis of historical price variability, the prices of the largest exporting countries are selected as world market prices. The largest exporting countries are determined based on FAOSTAT trade data from 1961 to 2006 (FAOSTAT, 2009). The US is selected as the world market price for wheat, France for barley, and Canada for rapeseed; Figure 4.6 shows their price data.

Figure 4.6: Nominal border prices (FOB) used for the calibration of the coefficient of variation of prices in ESIM



Source: own development based on data from Anderson and Valenzuela (2008)

In Figure 4.6, it can be observed how the linear trends are slightly positive due to low price levels before 1972; however, if only the period after is analyzed, no clear price tendency is identified. Therefore, the linear trends are not used for the determination of price uncertainty; for this purpose the price log returns is used. In finance, the expression return - in the context of log returns - is used to denominate the ratio or share of two price values in different periods. For example, P_t/P_{t-1}, where P_t is the price in period t and P_{t-1} the price in period $t-1$. The log returns refer to the natural logarithm of such a ratio, as follows:

$$ln\left(\frac{P_t}{P_{t-1}}\right). \quad (4.13)$$

This figure approximates the percentage difference between the prices in different periods of time and is often used in finance to determine volatility (Artavia et al., 2010).

The analysis of log returns produced the results presented in Table 4.1.

Table 4.1: Standard deviation of the log returns of the nominal border prices used for the calibration of the coefficient of variation of prices in ESIM

	SD of log returns
Wheat (USA)	25
Barley (France)	27
Rapeseed (Canada)	21

Source: own calculations based on data from Anderson and Valenzuela (2008)

The price variation in ROW generated with ESIM was calibrated to approximate the standard deviations given in Table 4.1. More specifically, the coefficient of variation of world markets generated with ESIM:

$$CV_{\text{row}} = \frac{SD_{row}}{\text{E}[P_{row}]} \quad (4.14)$$

– where, CV_{row} is the coefficient of variation, SD_{row} is the price standard deviation, and $\text{E}[P_{row}]$ is the price expected value for the ROW – is the figure calibrated to the values of Table 4.1.

Notice that the standard deviation of the log returns can be considered as an approximation of a coefficient of variation. Log returns approximate price

percentage deviations from an observation with respect to one period of time behind. If we consider the last period as the expected price value – which is a valid model for price data – then the standard deviation of the log returns is a coefficient of variation. Therefore, the calibration of the coefficient of variation obtained with repetitive model solves with the standard deviation of the log returns was considered as valid. However, it would have been more precise to compute simple returns:

$$\frac{P_t}{P_{t-1}} - 1, \tag{4.15}$$

instead of log returns, since log returns only approximate (not being exact) the percentage values. This detail was overseen while conducting the calibration and will be corrected in future research. Nonetheless, whether to use log or simple returns does not affect the evaluation of the accuracy of different quadrature formulas conducted in this study.

The calibration work is a process of repetitive manual manipulation of the supply elasticities until the desired price CV is achieved. For this exercise, a new parameter per commodity, ea (elasticity adjustment parameter), is created and the input and output price elasticites are adjusted by it. The ea is a factor applied to all elasticities of the respective commodity. The final values of the elasticity adjustment parameter are presented in Table 4.2.

Table 4.2: Final values of the elasticity adjustment parameter (*ea)* used as a factor to reduce the input and output price elasticities of supply

$ea_{for\ all\ commodities}$	0.300
ea_{cwheat}	0.150
ea_{barley}	0.200
ea_{rapeseed}	0.155

Source: own calculations

Table 4.2 shows that not all supply elasticities were reduced in the same proportion. First trials were run with such a formula, but the desired variance was not achieved. Thus, all supply elasticities were reduced to 30% of their original value and the specific elasticities for common wheat, barley and rapeseed were reduced to 15%, 20% and 15.5% of their original values, respectively.

In the deterministic version, own price and input price elasticities are very similar across countries, while cross price elasticities diverge to a higher degree. This is a result of the calibration of the system of elasticities to fulfill the conditions derived from economic theory theory, which are homogeneity of degree zero in input and output prices, symmetry of compensated substitution effects and non-negativity of the own price effect (Banse et al. 2007). In the calibration process, own and input price elasticities are held constant and only cross price elasticities are allowed to vary (Banse et al. 2007).

Microeconomic Foundations after New Calibration

After the adjustment of the elasticities for the stochastic version, the condition of homogeneity of degree zero is maintained since the ea from above is used for all the elasticities relevant to one equation. For example, if $\varepsilon_{1,1}$ is the own price elasticity, $\varepsilon_{1,2}, \cdots, \varepsilon_{1,n}$ are the cross price elasticities, and $\varepsilon_{1,int}, \varepsilon_{1,lab}, \varepsilon_{1,cap}$ are the elasticities with respect to input prices, the homogeneity condition states that:

$$\varepsilon_{1,1} + \varepsilon_{1,2} + \cdots + \varepsilon_{1,n} + \varepsilon_{1,int} + \varepsilon_{1,lab} + \varepsilon_{1,cap} = 0. \tag{4.16}$$

Then, by applying ea_1 as a factor to each elasticity:

$$ea_1\varepsilon_{1,1} + ea_1\varepsilon_{1,2} + \cdots + ea_1\varepsilon_{1,n} + ea_1\varepsilon_{1,int} + ea_1\varepsilon_{1,lab} + ea_1\varepsilon_{1,cap} = 0, \tag{4.17}$$

the condition homogeneity of degree zero is not altered.

The condition of convex profit functions with respect to product prices (non-negativity of own price elasticities), which applies under the assumption of profit maximizing behavior of farmers, is also still satisfied after the adjustment with ea.

If we let:

$$\boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_{1,1} & \varepsilon_{1,2} & \cdots & \varepsilon_{1,n} \\ \vdots & \vdots & \ddots & \vdots \\ \varepsilon_{n,1} & \varepsilon_{n,2} & \cdots & \varepsilon_{n,n} \end{bmatrix} \tag{4.18}$$

be the matrix of own and cross price elasticities, then, according to own deductions based on Jechlitschka et al. (2007), the desired curve properties are given if the leading principal minors of $\boldsymbol{\varepsilon}$ are greater than or equal to zero. Then, the multiplication of the elasticities with the respective ea does not alter the original fulfillment of this condition.

Finally, the symmetry condition which simplified to a two product model is given by (Jechlitschka et al., 2007):

(4.19) $\varepsilon_{1,2} p_1 q_1 = \varepsilon_{2,1} p_2 q_2$,

is the only condition which is no longer fulfilled after the adjustment of the elasticities. This occurs because the elasticity adjustment factors applied on the cross price elasticities of common wheat, barley and rapeseed are not the same, as can be observed in Table 4.2. This must be kept in mind when conducting the analysis of results.

4.4 Keeping the Number of Stochastic Variables Small

As already mentioned, the crops wheat, barley and rapeseed in all countries are considered random in the stochastic version of ESIM. The 3 stochastic crops times the 30 countries simulated in ESIM gives a total of 90 stochastic variables. If, for example, we use the Cartesian product of the Simpson's rule – which uses three points (see Chapter 2) –, then , we would need to run the model $3^{90} = 8.7 \times 10^{42}$ times to get an approximation of the distribution of model results of degree 2. There are, of course, more efficient formulas like those presented in Chapters 2 and 3. Nonetheless, it is desirable to keep the number of stochastic variables as small as possible without losing important sources of uncertainty, in order to keep the integration problem as simple as possible.

With this objective in mind, the number of stochastic variables in the EU was reduced based on EU shares of production per Member States, as well as on the correlation coefficients between stochastic variables. In countries with small shares of production, the yields are not shocked; highly correlated variables are grouped. The EU share of production is calculated using data from FAOSTAT for 1961-2006. For the comparison over that period of time, the average of production for every country is computed and these are put in relation to the average EU production. The correlation is computed for the calculated yield deviates for 1961-2006. For the countries put together, new deviates are computed based on the yield data for the region as a whole. With this analysis, the number of stochastic variables could be reduced to 42 (see Table 4.3). Grouped countries receive the same stochastic shock.

For computation of the yield deviates, FAOSTAT data (1961-2006) is used (FAOSTAT, 2010). In FAOSTAT, the yield data for Belgium and Luxembourg is given for both countries together from 1961 to 1999; but since 2000, these countries are treated separately. For Czech Republic and Slovakia, the data is given as one country until 1992 and since then as two separate regions. Therefore, Belgium-Luxembourg and Czech Republic-Slovakia are kept as

groups. For the EU New Member States (NMS), complete time series are available only for some of the countries (Poland, Czech Republic-Slovakia, Hungary, Romania and Bulgaria). For the Baltic States and Slovenia, the data is only available since 1992 (after the end of the communist era). For rapeseed, in Cyprus, Malta and Portugal no data is available (no production) and for Bulgaria, Greece and Spain the data is incomplete. Since the countries with incomplete data have small shares of production in the EU, yield uncertainty is not simulated for these.

Table 4.3 presents the results of the grouping analysis. The data is organized in order of the level of production, so that the largest producers are on the top and the smallest on the bottom. The analysis resulted in a total of 42 stochastic variables: 16 for wheat, 17 for barley and 9 for rapeseed. The mayor results for wheat are that Ireland is grouped with the UK; Portugal with Spain; and Denmark, Sweden and Finland are also put together. Ireland and Portugal represented each about 0.5% of total EU production and since the correlation and the similitude of standard deviation is high with their partner country they are grouped. Denmark, Sweden and Finland present EU production shares of approximately 2.2%, 1.5% and 0.5% and the standard deviations were 8% 10% and 15% respectively; the correlation between Denmark and Sweden was 63%, between Denmark and Finland 30%, and between Finland and Sweden 43%. It can be argued that Finland's data is not so similar to that of Denmark and Sweden; nonetheless, due to its relatively low level of production and the aim of reducing dimensionality, it was still put together with the other two countries.

The main results for barley are very similar to those for wheat, with the main difference being that Finland is not grouped with Denmark and Sweden. For rapeseed, the most important aspect is that the level of production of many countries is very low or none, which results in less stochastic variables than wheat and barley.

Table 4.3: Grouping of countries

	With stochastic variable			Without stochastic variable		
	Wheat	Barley	Rapeseed	Wheat	Barley	Rapeseed
1	France	Germany	Germany	Cyprus	Cyprus	Cyprus
2	Germany	France	France	Malta	Malta	Malta
3	UK, Ireland	UK, Ireland	Poland	Slovenia	Slovenia	Slovenia
4	Italy	Spain, Portugal	UK, Ireland	Baltic States	Baltic States	Baltic States
5	Poland	Denmark, Sweden	Czech Republic, Slovakia			Belgium
6	Romania, Bulgaria	Poland	Denmark, Sweden, Finland			Luxembourg
7	Spain, Portugal	Czech Republic, Slovakia	Hungary			Netherlands
8	Czech Republic, Slovakia	Romania, Bulgaria	Austria			Spain
9	Hungary	Finland	ROW*			Portugal
10	Denmark, Sweden, Finland	Austria				Italy
11	Greece	Hungary				Romania
12	Netherlands, Belgium, Luxembourg	Italy				Bulgaria
13	Austria	Netherlands, Belgium Luxembourg				Greece
14	Turkey	Greece				Turkey
15	US	Turkey				US
16	ROW*	US				
17		ROW*				

Source: Artavia et al. (2008)

4.5 The Deviates: Stationarity, Normality and Final Probability Density Function

After the grouping, the stochastic part of the yield time series is determined as the deviates from estimated linear trends, and the deviates are captured as shares. For example, let $y_{i,j}$ be any of the observed values for the variable i, where $i = (1,2,\cdots,n)$, in the year j, where $j = (1,2,\cdots,m)$; and let $\hat{y}_{i,j}$ be the estimated trend value for the same variable in the same year. Then the observed deviate $z_{i,j}$ is been captured by

$$z_{i,j} = \frac{y_{i,j}}{\hat{y}_{i,j}} - 1. \tag{4.20}$$

Proceeding in this way for all variables and all years, the matrix of deviates $\mathbf{Z}_{n\times m}$ is computed. The final matrix of deviates is arranged as follows:

$$\mathbf{Z}_{42\times 46} = \begin{bmatrix} z_{FR.w,1} & z_{FR.w,2} & \cdots & z_{FR.w,m} \\ z_{GE.w,1} & z_{GE.w,2} & \cdots & z_{GE.w,m} \\ \vdots & \vdots & & \vdots \\ z_{ROW.w,1} & z_{ROW.w,2} & \cdots & z_{ROW.w,m} \\ z_{FR.b,1} & z_{FR.b,2} & \cdots & z_{FR.b,m} \\ z_{GE.b,1} & z_{GE.b,2} & \cdots & z_{GE.b,m} \\ \vdots & \vdots & & \vdots \\ z_{ROW.b,1} & z_{ROW.b,2} & \cdots & z_{ROW.b,m} \\ z_{FR.r,1} & z_{FR.r,2} & \cdots & z_{FR.r,m} \\ z_{GE.r,1} & z_{GE.r,2} & \cdots & z_{GE.r,m} \\ \vdots & \vdots & & \vdots \\ z_{ROW.r,1} & z_{ROW.r,2} & \cdots & z_{ROW.r,m} \end{bmatrix}, \tag{4.21}$$

for $n = 42$ variables and $m = 46$ years.

In the elements of $\mathbf{Z}$, the first index indicates the country and the crop at the same time and the second index determines the observation year. For example, $FR.w,1$ is the value for wheat in France in year number one. Then, it can be observed in (4.21) that the variables of wheat were located in the upper block of the matrix, those of barley in the middle block (indices with: $country.b$) and rapeseed in the bottom block (indices with: $country.r$).

Since the deviates are calculated in relation to a linear trend estimated by the method of ordinary least squares (OLS), the expected value of every stochastic variable is assumed to be zero. Its determination can be expressed by:

$$\mathrm{E}[\boldsymbol{z}] = \frac{1}{m}(z_1 + \cdots + z_{\mathrm{m}}) = \mathbf{0}. \tag{4.22}$$

The standard deviations of the stochastic variables are presented in Table 4.4. This table also shows the results on the tests for stationarity and normal distribution of the variables.

Table 4.4: Standard deviation (in %), regular Dickey-Fuller Test (95%)* and tests of normal distribution (95%)** of the yield deviates

Wheat				Barley				Rapeseed			
	SD	DF	ND		SD	DF	ND		SD	DF	ND
FR	0.08	Yes	Yes	FR	0.08	Yes	Yes	FR	0.12	Yes	Yes
GE	0.07	Yes	Yes	GE	0.07	Yes	Yes	GE	0.10	Yes	Yes
UK-IR	0.09	Yes	Yes	UK-IR	0.06	Yes	Yes	UK-IR	0.14	Yes	Yes
PL	0.11	Yes	No	PL	0.13	Yes	Yes	PL	0.17	Yes	Yes
IT	0.08	Yes	Yes	IT	0.13	No	No	DK-SW-FI	0.10	Yes	Yes
ES-PT	0.16	Yes	Yes	ES-PT	0.19	Yes	Yes	AT	0.12	Yes	Yes
DK-SW-FI	0.09	Yes	Yes	FI	0.14	Yes	Yes	CZ-SK	0.16	Yes	Yes
BE-NL	0.09	Yes	Yes	DK-SW	0.09	Yes	No	HU	0.17	Yes	Yes
GR	0.17	Yes	Yes	BE-NL	0.09	Yes	Yes	ROW	0.07	Yes	Yes
AT	0.10	Yes	Yes	GR	0.15	Yes	Yes				
CZ-SK	0.15	No	Yes	AT	0.10	Yes	Yes				
HU	0.23	No	Yes	CZ-SK	0.17	No	Yes				
RO-BG	0.21	Yes	Yes	HU	0.20	Yes	Yes				
TU	0.11	Yes	No	RO-BG	0.21	Yes	Yes				
US	0.07	Yes	Yes	TU	0.10	Yes	No				
ROW	0.06	Yes	Yes	US	0.08	Yes	No				
				ROW	0.09	Yes	Yes				

* Test of rejection of the H0; if we cannot reject it, it indicates that the process has a unit root.

** If one of 4 tests (Chi-Squared, Shapiro Wilks, Anderson-Darling, and Cramer-von Mises) rejects the H0 of Normal Distribution it was marked with "No" in the table.

SD: standard deviation; DF: regular Dickey-Fuller Test; ND: normal distribution.

AT: Austria, BE: Belgium-Luxembourg, BG: Bulgaria, CY: Cyprus, CZ: Czech Republic, DK: Denmark, EE: Estonia, FI: Finland, FR: France, GE: Germany, GR: Greece, HU: Hungary, IE: Ireland, IT: Italy, LV: Latvia, LT: Lithuania, MT: Malta, NL: Netherlands, PL: Poland, PT: Portugal, RO: Romania, ROW: Rest of the World, SK: Slovak Republic, SI: Slovenia, ES: Spain, SW: Sweden, TU: Turkey, UK: United Kingdom, US: United States of America

Source: own calculations

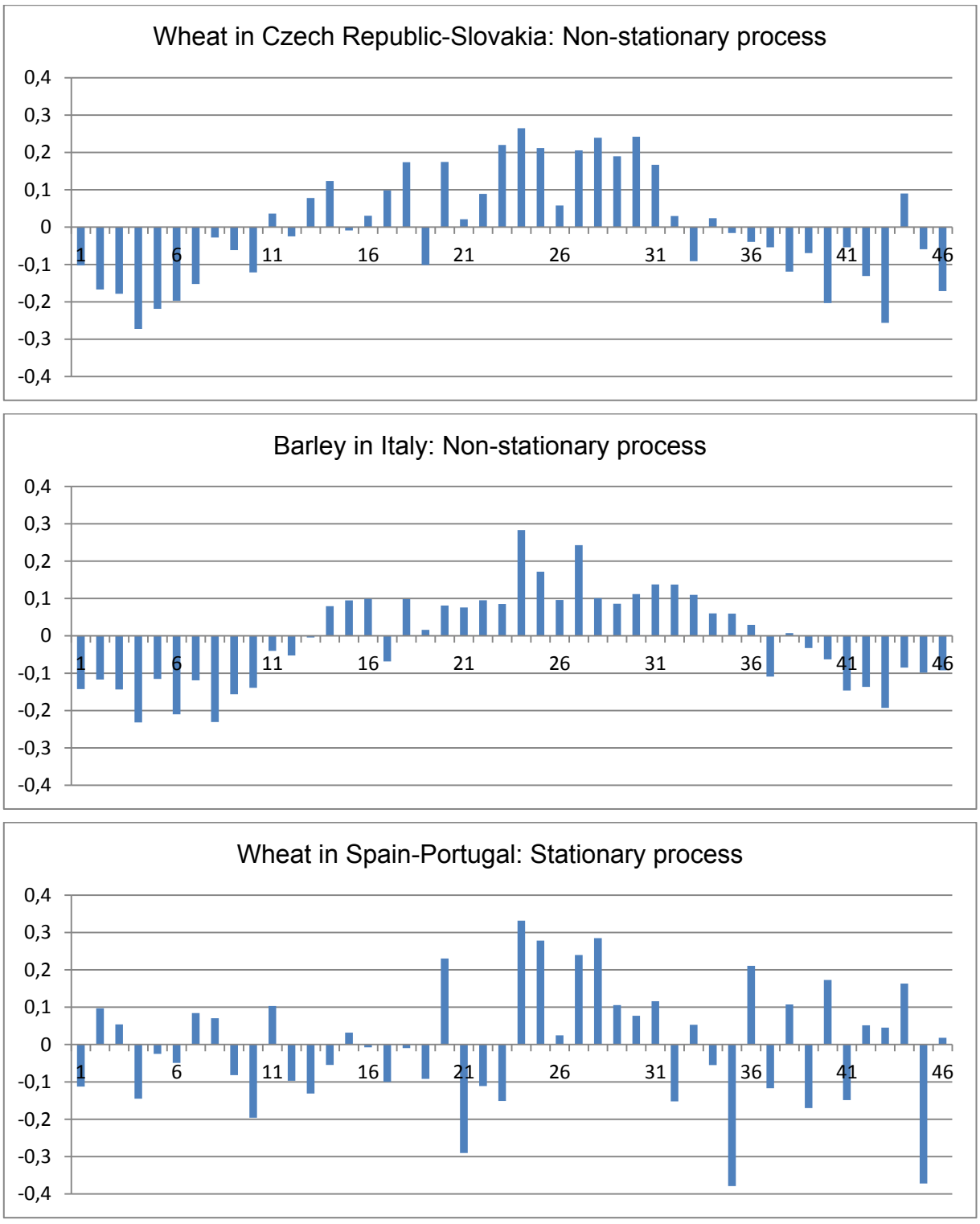
Based on Table 4.4, the EU can be roughly classified in three regions corresponding to their level of yield variability: a first group of countries of western Europe (Germany, France, UK-Ireland, Belgium-Luxembourg, Denmark-Sweden, and Austria) with relatively low variation of yields (around 10% or less), a second group of NMS and south EU countries (Czech Republic, Hungary, Romania-Bulgaria, Greece, and Spain-Portugal) with relatively high standard deviation (more than 15%), and a last group of countries (Poland, Italy, Turkey, and Finland) of middle variance with a standard deviation between 10% and 15%. Note that the variance for rapeseed is in general higher. Also, note that the ROW presents almost the lowest variability between all countries for all three commodities. This occurs because the ROW is a big country aggregate where positive and negative weather conditions occur simultaneously and neutralize each other.

With respect to the test for stationarity, the regular Dickey-Fuller test is carried out. It is tested whether the expected value of the time series remains constant and whether the variance does not depend on time (Gujarati, 2003). In Table 4.4, it can be observed that for some few stochastic variables, the hypothesis null of a process with a unit root cannot be rejected; this is the case of wheat in the group Czech Republic-Slovakia and in Hungary, and the case of barley in Italy and in Czech Republic-Slovakia.

With respect to the proof of normal distribution, four tests are applied. If with one of those tests the H0 of normal distribution is rejected, then it is indicated as not normally distributed in Table 4.4. The following stochastic variables were not proved as normally distributed: for wheat: Poland and Turkey; for barley: Italy, Denmark-Sweden, Turkey and the US; for rapeseed: none.

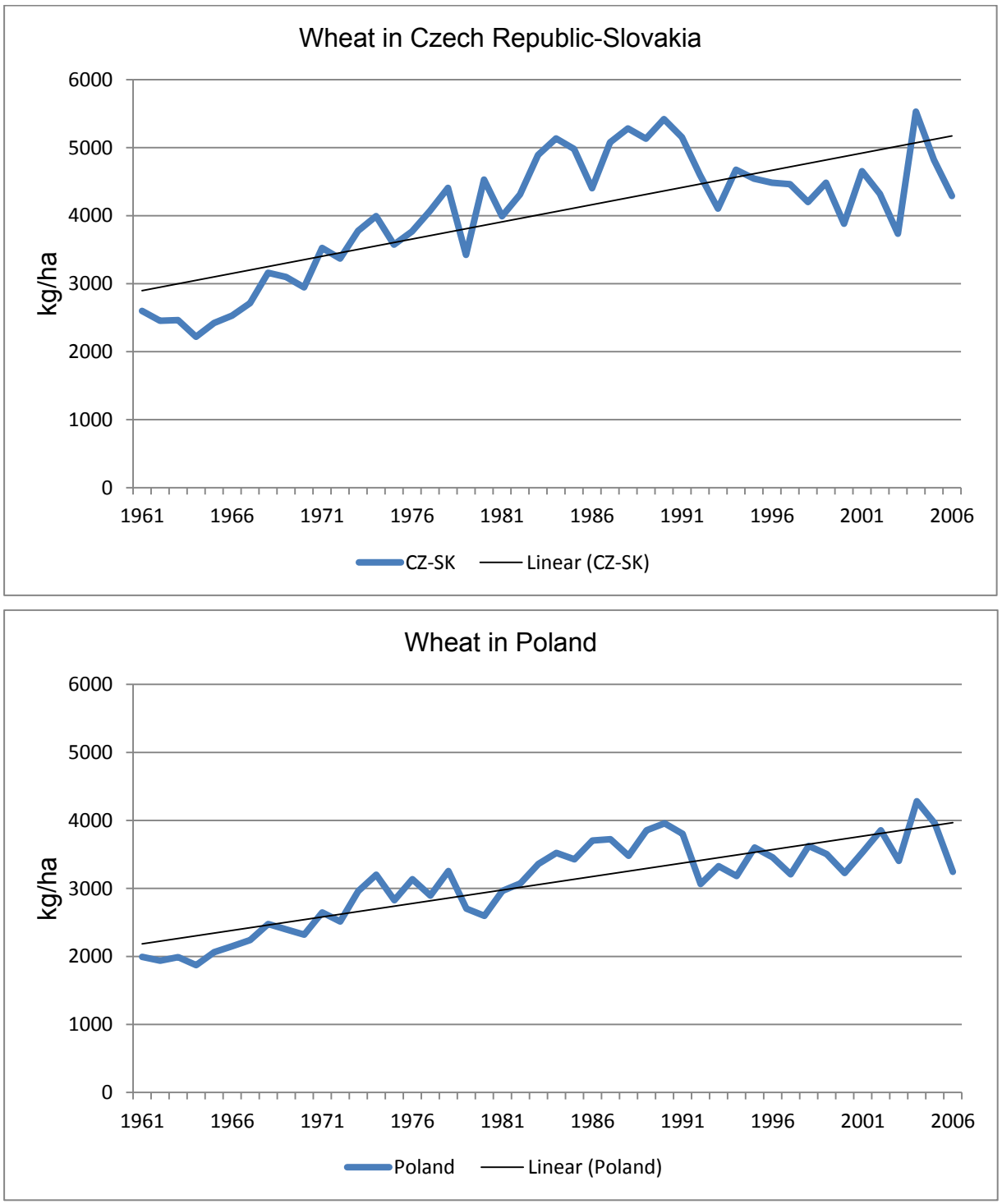
The main reason for failed stationarity and normality tests is that no structural changes are considered in the trends. Figures 4.7 to 4.9 show the problems generated with this procedure. Figure 4.7 illustrates two non-stationary and one stationary process, Figure 4.8 shows structural breaks on productivity due to the transformation to market economies of some of the NMS, and Figure 4.9 exposes the slowdown in productivity growth observed in many industrialized countries since approximately 1990.

Figure 4.7: Stochastic variables with stationary and non-stationary processes



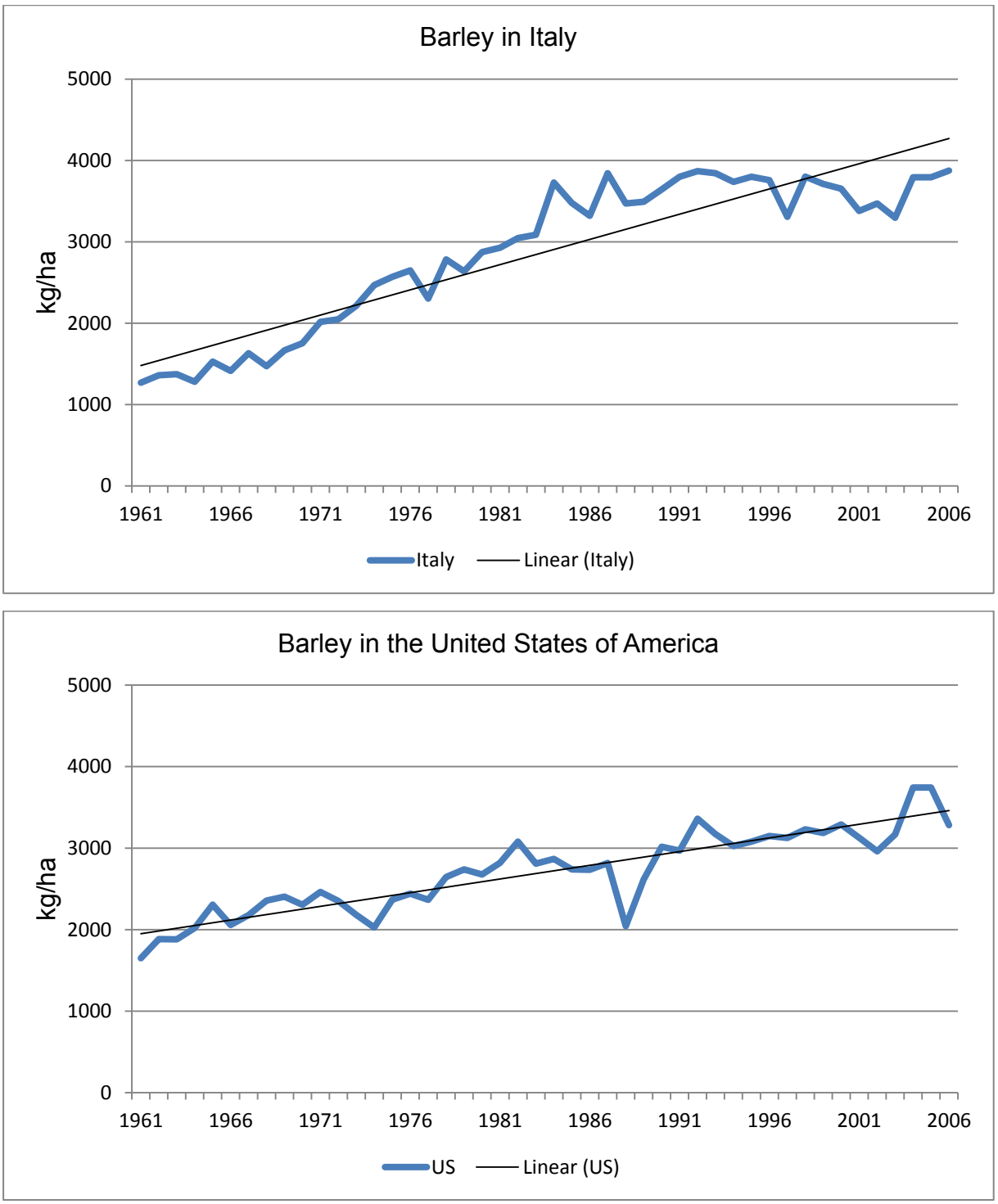
Source: own development

Figure 4.8: Yield time series and linear trend for wheat in some EU New Member States (NMS)



Source: own development based on FAOSTAT data (1961-2006) for production, area and yield (FAOSTAT, 2010)

Figure 4.9: Yield time series and linear trend in some industrialized countries



Source: own development based on FAOSTAT data (1961-2006) (FAOSTAT, 2010)

In Figure 4.8, it can be observed that since 1992 Czech Republic-Slovakia and Poland – indeed all the NMS apart from Cyprus and Malta – suffered a significant productivity fall, which was associated with the disruptions that the organizational systems suffered with the abrupt transition from planned to market economies. It can be seen that in this way, the linear trend fails to capture the annual uncertainty attached to yields, for instance, the weather induced variability.

In the cases of barley in Italy and the US (see Figure 4.9), a decrease in productivity growth is detected since approximately the mid-1980s for Italy and perhaps since the beginning of the 1990s for the US. This is a phenomenon which can be observed in the data in almost all industrialized countries and again, only one linear trend fails to capture the annual yield variability well.

Note that for the stochastic variable for barley in the US, only in one test (Shapiro-Wilks) the H0 is rejected. Indeed, its decrease in productivity growth is not strong. Nonetheless, the data series presents several downward picks of significant magnitude in 1973, 1988, and 2002, with the one from 1988 being the most pronounced. This probably is the reason for the rejection of H0 with the Shapiro-Wilks test. Such behaviors support the opinion of some researchers that yield deviates may not always be normally distributed, since the negative consequences of extreme weather events may be stronger than the positive ones.

Even though some stochastic variables are not tested as stationary and others are not tested as normally distributed, for the generation of quadratures for this study, all stochastic variables are assumed to be normally distributed: $N(\mathbf{0}, \boldsymbol{\Sigma}[\boldsymbol{z}])$.

Also, take into consideration that the errors in the calculation of the deviates result in a wrong perception of their PDF. For example, the variance may be overestimated. However, as explained in Section 5.3, the PDF of the stochastic variables attached to the yield or supply equations of small countries in ESIM have a small effect on the EV and CV of world market prices.

5 Does the Rotation of Stroud's Octahedron Matter?

This chapter studies the accuracy of different quadratures formulas based on Stroud's n-octahedron[7]. For the analysis, the stochastic version of ESIM is used. First, the true value of model results is estimated using LHS quadratures with progressively increasing N. Then, the different quadrature formulas tested are described. Finally, the results obtained are presented and analyzed.

5.1 Approximation of the True Value of ESIM Results

As just mentioned, in Chapter 5 we determine the accuracy of approximation of the EV and CV of model results of different quadrature formulas. Equation (2.15) from Section 2.1.1 shows how to calculate the approximation error. However, the true value of the multivariate integration problems in ESIM, $\int_{\mathbb{R}^n} g(\mathbf{x})f(\mathbf{x})d\mathbf{x}$, cannot be computed since we do not have a polynomial which describes the relationships between the dependent variables (e.g. prices) and the stochastic variables. Furthermore, the equilibrium conditions in the ESIM, the embedded system of equations, and the high dimensionality make the determination of the true value of the model results an impossible task. Consequently, the true value must be approximated by determining the convergence point of the model results. Convergence is obtained by running the stochastic version of ESIM with LHS samples of progressively increasing size.

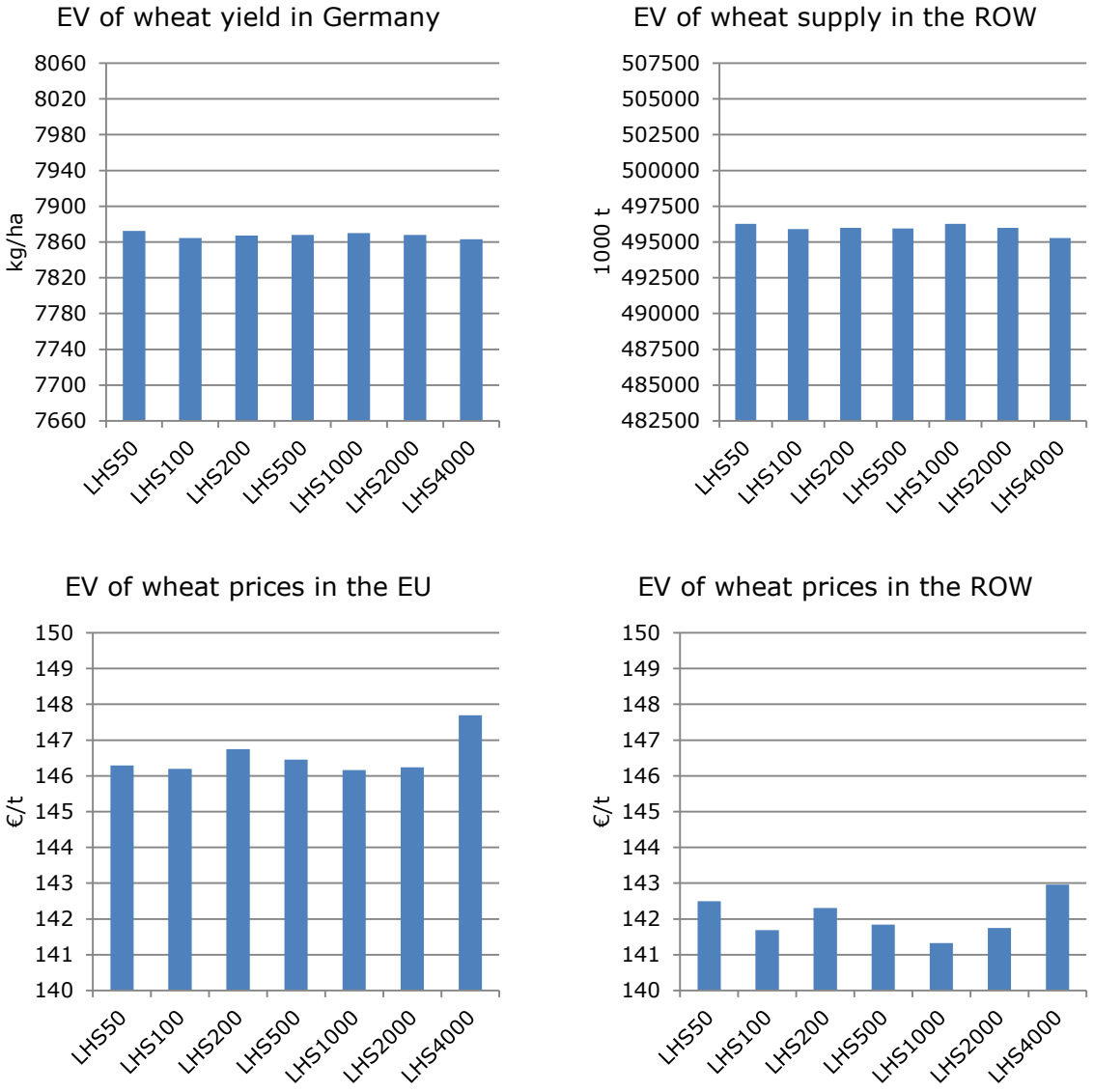
The LHS method offers the advantage that the smoothness of $g(\mathbf{x})$ is not essential for the accuracy of approximation of the method (see Section 2.2.2 for a detailed description of the LHS). However, when compared to numerical integration methods based on interpolation, it requires a much higher N to obtain quadratures of the same degree of precision. The procedure for the generation of the LHS quadratures is the one presented in Section 3.3 and can be summarized

[7] Please note that the results presented in Chapter 5 contain a small error in the simulation of export subsidies for wheat. In the stochastic version of ESIM, in some of the cases world market price levels are above the EU threshold price and export subsidies should not be applied. In Chapter 5 this detail was not considered and in those cases an export subsidy of 5 €/t is simulated. This mistake is not corrected here since it is not relevant for the analysis of the accuracy of the quadratures.

as: a) simulation of the standard normal deviates, and b) induction of the desired covariance matrix to the sample. The LHS samples are generated via Cholesky decomposition, as explained in Section 3.2, and the sample sizes used are: $N = 50$, $N = 100$, $N = 200$, $N = 500$, $N = 1000$, $N = 2000$ and $N = 4000$.

Figures 5.1 to 5.3 show a selection of the results obtained. The results are presented mainly for the ROW, as this is the largest country in ESIM and has the greatest influence in the determination of world market prices. The following order is followed for the presentation of results: i) the EV of supply and prices of wheat, b) the CV of supply of wheat, barley and rapeseed, and c) the CV of prices of wheat, barley and rapeseed. In some cases, the results for one of the EU Member States are presented in order to show whether there are differences with respect to the ROW, which may occur due to the simulated EU price polices. In these cases, the results for yield and prices in Germany are shown, since Germany is a large producer of all three products in the EU. Wheat is the only commodity for which the price policies are triggered with the stochastic runs, and, thus, special attention is paid to it. The convergence of the EV of model results is obtained easier than for the CV; thus, for the EV, only results for wheat are presented. Note that the units of the vertical axes in Figures 5.1 to 5.3 are set to 0.5%, for the purpose of comparing the results obtained for the different commodities. In Figure 5.1, the vertical axes are divided into units close to 0.5% of the results obtained with the LHS4000 sample.

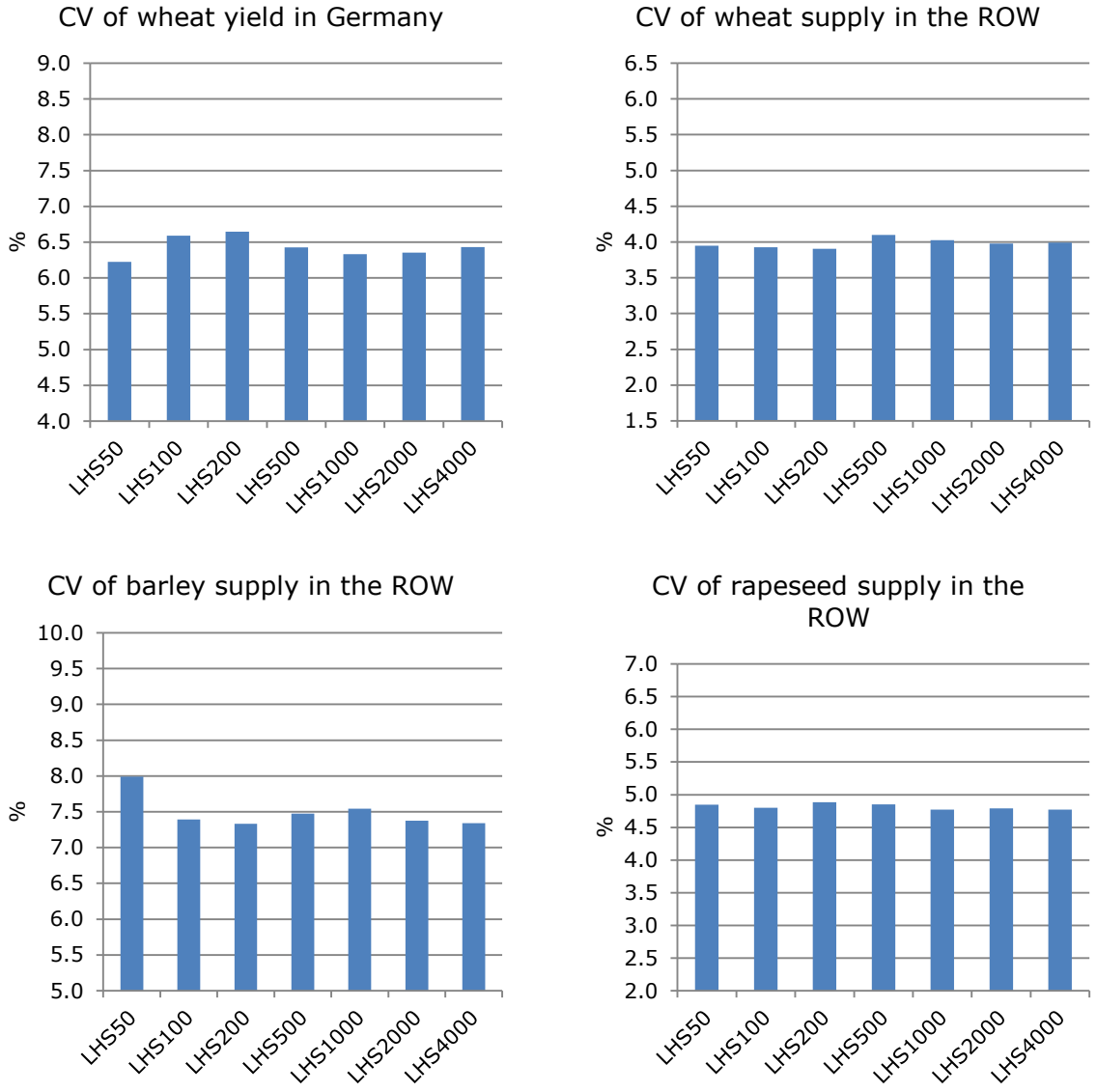
Figure 5.1 Expected value (EV) of wheat yield in Germany, wheat supply in the ROW, and wheat prices in the EU and in the ROW obtained with LHS quadratures of different sizes



Note: LHS50 – LHS4000 are the 7 samples of different *N* size (50, 100, 200, 500, 1000, 2000, and 4000) of the multivariate stochastic distribution obtained through Latin Hypercube Sampling (LHS).

Source: own calculations

Figure 5.2 Coefficient of variation (CV) of wheat yield in Germany and of the supply of wheat, barley and rapeseed in the ROW obtained with LHS quadratures of different sizes

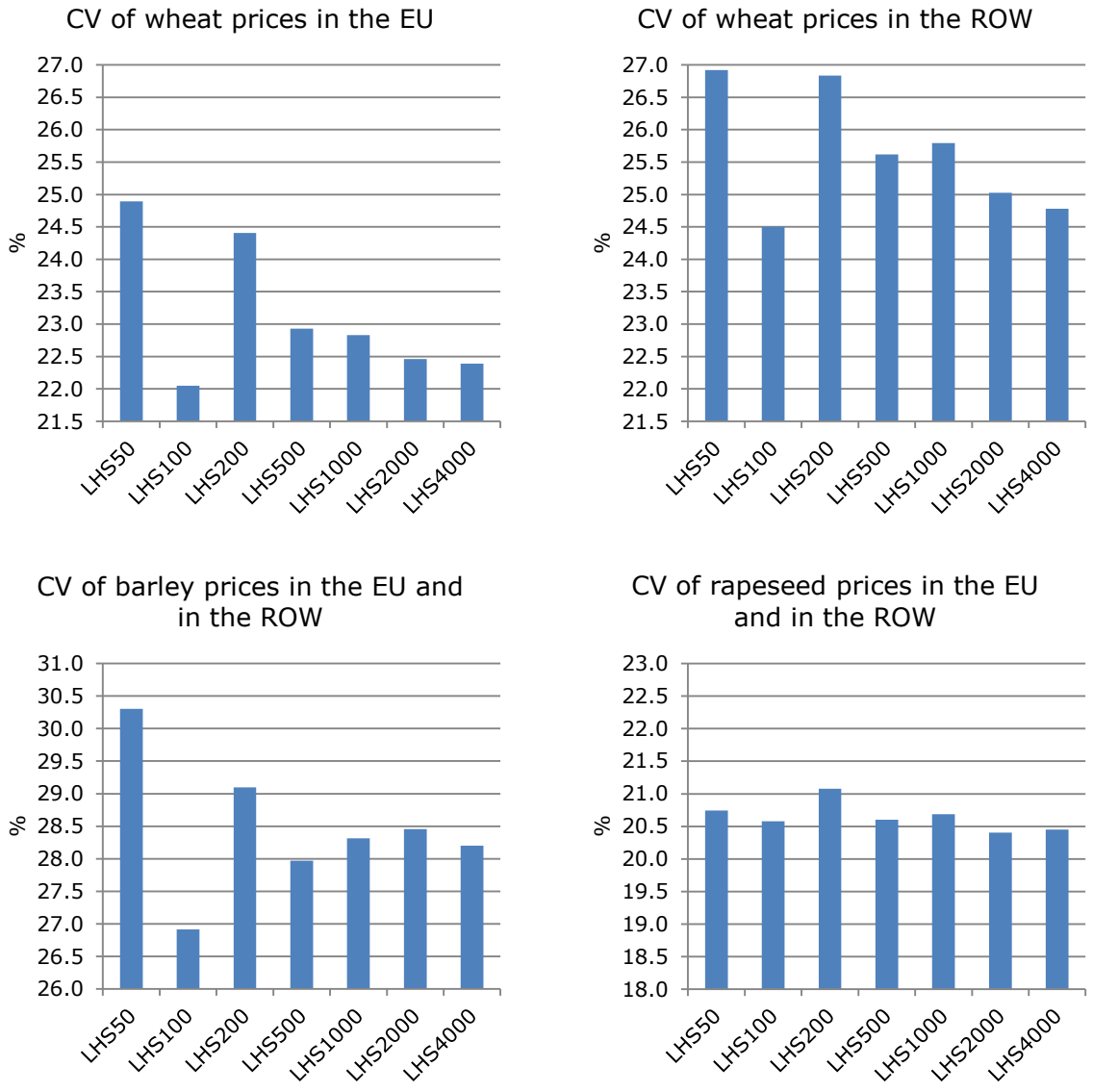


Note: LHS50 – LHS4000 are the 7 samples of different *N* size (50, 100, 200, 500, 1000, 2000, and 4000) of the multivariate stochastic distribution obtained through Latin Hypercube Sampling (LHS).

Note 2: The CV is computed using the corresponding EV for each sample size.

Source: own calculations

Figure 5.3 Coefficient of variation (CV) of wheat prices in the EU and of the prices of wheat, barley and rapeseed in the ROW obtained with LHS quadratures of different sizes



Note: LHS50 – LHS4000 are the 7 samples of different *N* size (50, 100, 200, 500, 1000, 2000, and 4000) of the multivariate stochastic distribution obtained through Latin Hypercube Sampling (LHS).

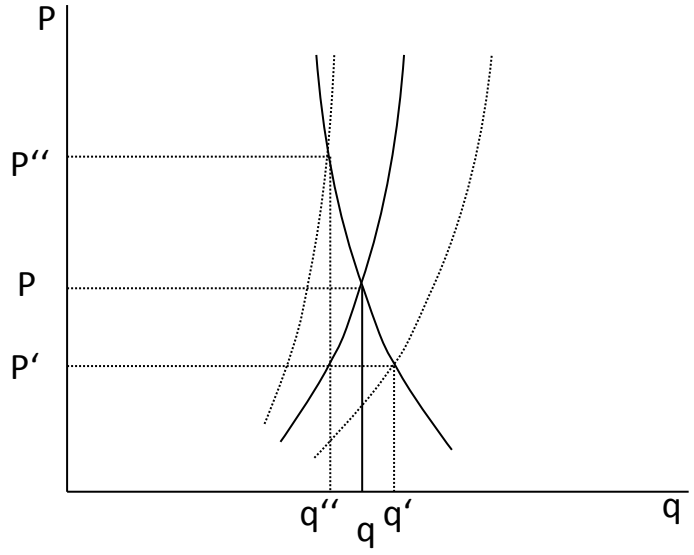
Note 2: The CV is computed using the corresponding EV for each sample size.

Source: own calculations

In Figures 5.1 to 5.3, it can be observed that the convergence of the EV and of the CV of the yield and supply results obtained with the LHS samples of increasing size is easier to achieve than the convergence of the CV of the commodity prices. Thus, the convergence of the CV of the world market prices serves as the reference for the identification of the true value.

One of the reasons why it is easier to obtain convergence of the supply results is that for the stochastic version yield and supply elasticities with respect to prices have been reduced. This was done in order to simulate the short term supply reactions to the yield shocks (see Table 4.2). Accordingly, the supply shocks introduced result in large price effects (see Figure 5.4). Due to the small elasticities, yield and supply react little to prices changes.

Figure 5.4 The effect on prices of inelastic isoelastic supply curves



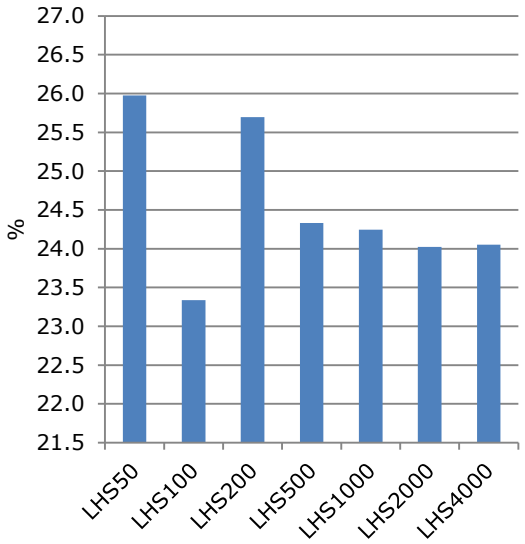
Source: own development

Figure 5.3 shows that convergence differs among crops. Convergence of the CV of wheat prices is the most difficult to achieve, while convergence of the CV of rapeseed prices is the easiest. The results for LHS1000, LHS2000 and LHS4000 appear to converge for barley and rapeseed. However, for wheat the difference of the results with the samples of size 1000 and 4000 is 1% and it is not clear whether convergence is actually achieved. Since it is assumed that the accuracy of approximation of model results increases with increasing sample size, the results obtained with LHS4000 are used as the approximated true value.

The lesser convergence for wheat price results may be a consequence of the complex price policies simulated for it; namely: intervention price, export subsidies, export subsidy limit, TRQ, and threshold price, which are triggered depending on the net trade position of the EU. For barley, price policies are also

simulated; however, these are not triggered, since in all runs the EU is in the net trade position where the EU domestic price equals the world market price. To test the effect of the price policies simulated for the wheat market, ESIM was run again with LHS50 – LHS4000, but under the assumption of a liberalized EU wheat market. The results are presented in Figure 5.5.

Figure 5.5 Coefficient of variation (CV) of wheat prices in the ROW obtained with LHS quadratures of different sizes under the assumption of a liberalized EU wheat market



Note: LHS50 – LHS4000 are the 7 samples of different *N* size (50, 100, 200, 500, 1000, 2000, and 4000) of the multivariate stochastic distribution obtained through Latin Hypercube Sampling (LHS).

Note 2: The CV is computed using the corresponding EV for each sample size.

Source: own calculations

By comparing the wheat prices in the ROW in Figure 5.3 with Figure 5.5 it can be seen that convergence is achieved faster in the liberalized market. This indicates that the simulated price policies add complexity to the numerical integration problem. For example, the policies result in threshold points where price levels change abruptly (see Figures 4.1-4.4), making the approximation of model results more difficult.

In Figures 5.3 and 5.5, it can also be seen that it is more difficult to approximate the CV of the prices of wheat (even liberalized) and barley than rapeseed. Among other reasons, this behavior may occur because, in ESIM, soft wheat and barley have stronger cross relationships with each other than with rapeseed. Thus, stochastic variables (dimensions) with a significant effect on wheat and barley markets are larger than in the case of the rapeseed market, adding further complexity to the numerical integration to the prices of those markets.

5.2 The Different Quadrature Formulas Tested

The accuracy analysis comprised eight quadrature formulas all based on Stroud's theorem from 1957 (see Section 3.1.1). The approach used for the generation of the quadratures for the multivariate normal PDF of the stochastic space, $\mathbf{x}\sim N(\mathbf{0},\boldsymbol{\Sigma})$, is, as suggested in Chapter 3, to first approximate the multivariate standard normal PDF, $\mathbf{x}\sim N(\mathbf{0},\mathbf{I_n})$, and then to induce the desired covariance matrix using one of many possible methods.

For the approximation of $\mathbf{x}\sim N(\mathbf{0},\mathbf{I_n})$, we use: i) the formula proposed by Artavia et al., which is the Q^n with its vertices lying on the coordinate axes, and ii) Arndt's formula, which is the Q^n rotated in such a way as to benefit from the space in direction of the vertices of a C^n with vertices $(\pm a,\pm a,\cdots,\pm a)$ whose centroid coincide with the centroid of the region. Remember that the Q^n from Arndt gives the same rotation as Stroud's Q^n for the C^n, which has the purpose of generating points which are interior to the region for all n; this explains the rotation of Stroud's Q^n (see Section 3.1).

In addition to the different formulas to get the Q^n, it was studied whether the arrangement of the n-coordinate system (the arrangement of the stochastic variables in $\mathbf{Z}$) has consequences for the accuracy of approximation of the model results. In order to evaluate this, the transformed octahedron, $Q_t^n = \mathbf{A\Gamma}$, is generated using the matrix of historical data $\mathbf{Z}$ with two different arrangements ($A1$ and $A2$). In this way, two different matrices, $\mathbf{A}_{\mathrm{A1}}$ and $\mathbf{A}_{\mathrm{A2}}$, are obtained from the factorization of the covariance matrices $\boldsymbol{\Sigma}[\boldsymbol{z}_{\mathrm{A1}}]$ and $\boldsymbol{\Sigma}[\boldsymbol{z}_{\mathrm{A2}}]$ in the form $\boldsymbol{\Sigma}[\boldsymbol{z}] = \boldsymbol{A}\boldsymbol{A}^T$.

In $\mathbf{Z}_{\mathrm{A1}}$, the coordinates are arranged in the same order as in the column for wheat in the grouping analysis (see Table 4.3) and as in (4.21). Figure 5.6 exemplifies the arrangement in $\mathbf{Z}_{\mathrm{A1}}$.

Figure 5.6 The arrangement of the n-coordinate system in the matrix $\mathbf{Z}_{\mathrm{A1}}$ of size $n \times m$

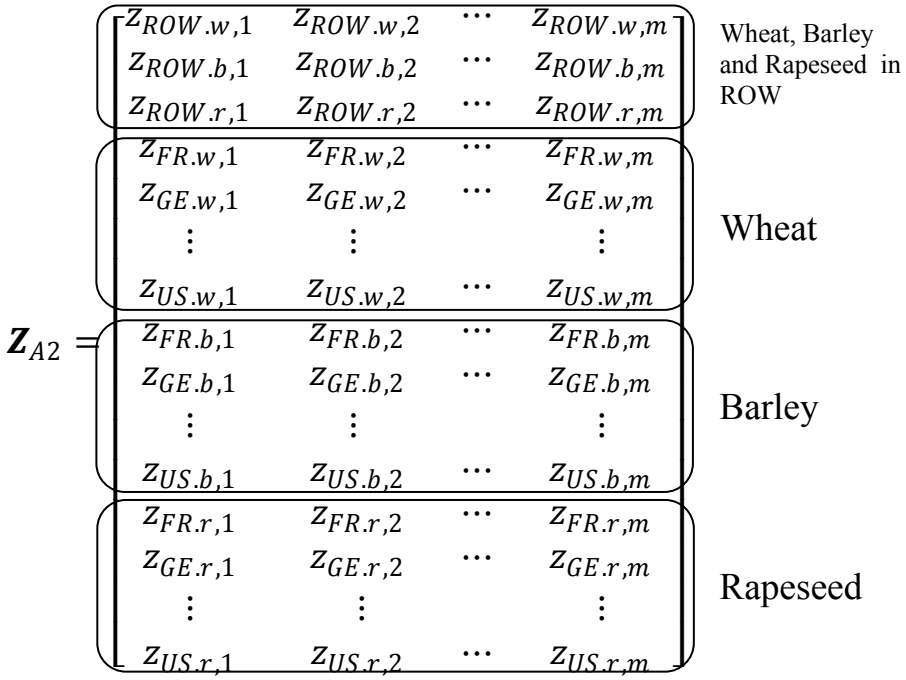
$$\boldsymbol{Z}_{A1} = \begin{bmatrix} z_{FR.w,1} & z_{FR.w,2} & \cdots & z_{FR.w,m} \\ z_{GE.w,1} & z_{GE.w,2} & \cdots & z_{GE.w,m} \\ \vdots & \vdots & & \vdots \\ z_{ROW.w,1} & z_{ROW.w,2} & \cdots & z_{ROW.w,m} \\ z_{FR.b,1} & z_{FR.b,2} & \cdots & z_{FR.b,m} \\ z_{GE.b,1} & z_{GE.b,2} & \cdots & z_{GE.b,m} \\ \vdots & \vdots & & \vdots \\ z_{ROW.b,1} & z_{ROW.b,2} & \cdots & z_{ROW.b,m} \\ z_{FR.r,1} & z_{FR.r,2} & \cdots & z_{FR.r,m} \\ z_{GE.r,1} & z_{GE.r,2} & \cdots & z_{GE.r,m} \\ \vdots & \vdots & & \vdots \\ z_{ROW.r,1} & z_{ROW.r,2} & \cdots & z_{ROW.r,m} \end{bmatrix} \begin{matrix} \\ \\ \\ \\ \text{wheat} \\ \\ \\ \\ \text{barley} \\ \\ \\ \\ \text{rapeseed} \end{matrix}$$

Note: The indices in the figure have the following meaning: FR.w, FR.b, FR.r: wheat, barley and rapeseed in France; GE.w, GE.b, GE.r: wheat, barley and rapeseed in Germany, ROW.w, ROW.b, ROW.r: wheat, barley and rapeseed in the ROW

Source: own development

In $\mathbf{Z}_{\mathrm{A2}}$, the stochastic variables $\boldsymbol{z}_{ROW.w}$, $\boldsymbol{z}_{ROW.b}$, and $\boldsymbol{z}_{ROW.r}$, which are important for the price determination in ESIM, are ordered at the beginning. Figure 5.6 represents the arrangement in $\boldsymbol{Z}_{\mathrm{A2}}$.

Figure 5.7 The arrangement of the n-coordinate system in $\mathbf{Z}_{\mathrm{A2}}$ of size $n \times m$



Note: The indices in the figure have the following meaning: FR.w, FR.b, FR.r: wheat, barley and rapeseed in France; GE.w, GE.b, GE.r: wheat, barley and rapeseed in Germany, ROW.w, ROW.b, ROW.r: wheat, barley and rapeseed in the ROW, US.w, US.b, US.r: wheat, barley and rapeseed in the US

Source: own development

In order to induce the desired variance and covariance of the stochastic variables in $\mathbf{Z}_{\mathrm{A1}}$ and $\mathbf{Z}_{\mathrm{A2}}$ to the quadratures $\boldsymbol{\Gamma}$ for the multivariate standard normal PDF, two linear transformation methods are tested: i) the one applying the Cholesky decomposition of $\boldsymbol{\Sigma}[\boldsymbol{z}_{\mathrm{A1}}]$ and $\boldsymbol{\Sigma}[\boldsymbol{z}_{\mathrm{A2}}]$, and ii) the one using the diagonalization of $\boldsymbol{\Sigma}[\boldsymbol{z}_{\mathrm{A1}}]$ and $\boldsymbol{\Sigma}[\boldsymbol{z}_{\mathrm{A2}}]$ (see Box 3.6).

By combining the quadrature formulas from Artavia et al. and Arndt, the arrangements of the n-coordinate system A1 and A2, and the methods to induce the covariance matrix (via Cholesky decomposition and the diagonalization method), the eight different quadrature formulas are obtained (see Table 4.5).

Table 5.1 The different quadrature formulas tested

Art.-A1-C	Str.-A1-C
Art.-A1-D	Str.-A1-D
Art.-A2-C	Str.-A2-C
Art.-A2-D	Str.-A2-D

Note: The quadrature formulas names mean the following: Art.: Artavia et al.'s formula, Str.: Arndt's formula with the rotation from Stroud's Q^n, A1 and A2: Arrangements 1 and 2 of the n-coordinate system, C and D: the linear transformations using the Cholesky decomposition (C) and the diagonalization method (D)

Source: own development

Note that the quadratures using Arndt's formula have been named "Str." for the purpose of avoiding confusions between "Art." and "Arn." in the presentation of the results. Remember that Arndt's Q^n presents the same rotation as Stroud's Q^n for the C^n with vertices $(\pm 1, \pm 1, \cdots, \pm 1)$. For this reason, Arndt is abbreviated with "Str.".

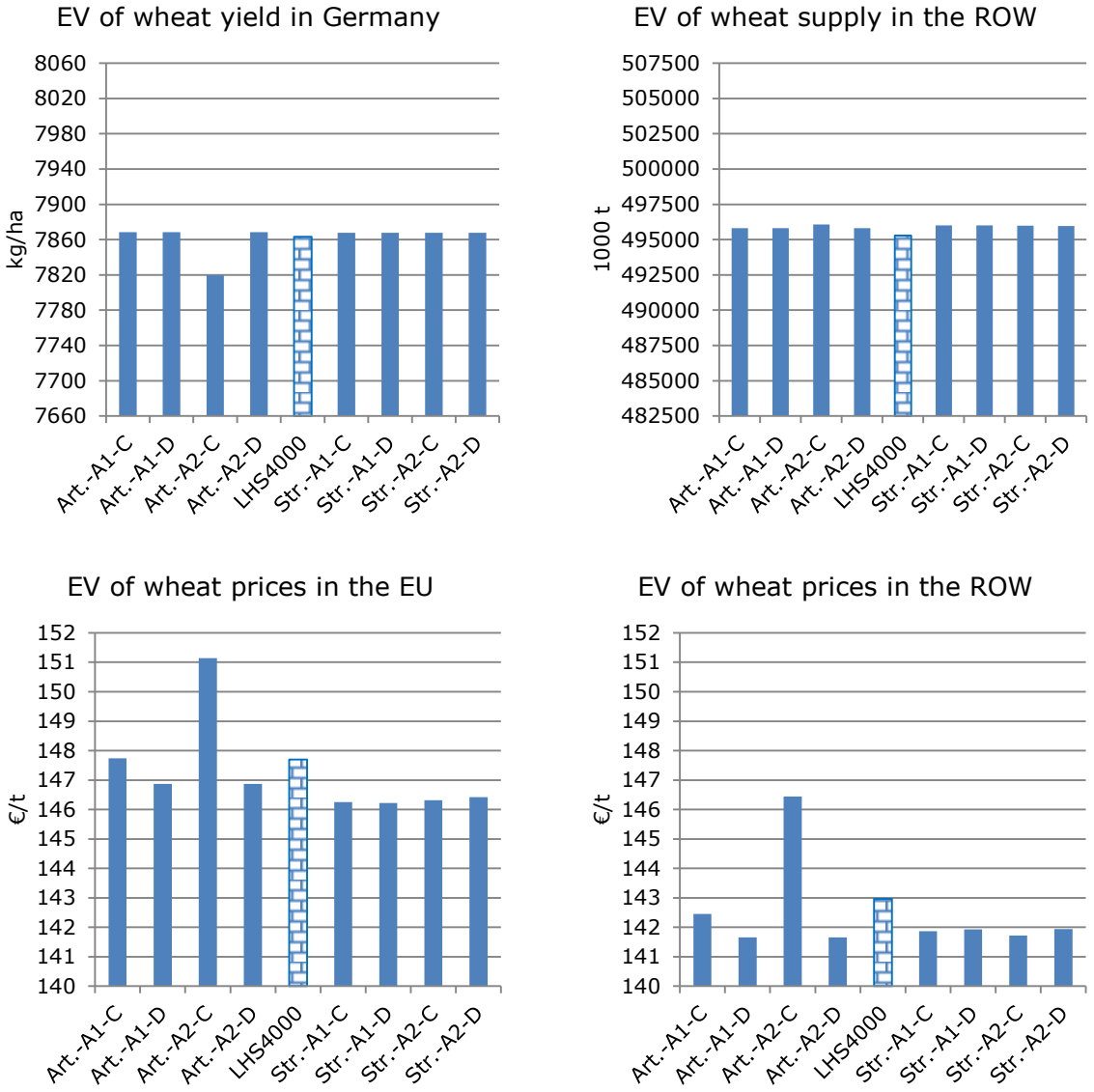
Note that in a similar way as the different quadrature formulas from Arndt and from Artavia et al. result in different rotations of the Q^n, the different quadrature formulas from Table 5.1 result in different rotations from the Q_t^n. Arrangements A1 and A2, and the methods to induce the covariance matrix result in different linear transformations of the Q^n obtained with Artavia et al.'s formula and with Arndt's formula.

5.3 Accuracy of Approximation of the Different Quadratures

5.3.1 Presentation of Results

For the determination of the approximation error of the different quadrature formulas (rotations of the Q_t^n) described above, ESIM is run using the different quadratures, and the obtained model results are evaluated against the approximated true value obtained with the LHS4000 sample. The same selection and sequence of ESIM results as in Section 5.1 is chosen, namely: a) the EV of supply and prices of wheat, b) the CV of supply of all three stochastic commodities, and c) the CV of prices of all three commodities. Also, the units of the vertical axes in Figures 5.8 to 5.10 are set to 0.5%, for the purpose of comparing the results obtained for the different commodities. In Figure 5.8, the vertical axes are divided into units close to 0.5% of the results obtained with the LHS4000 sample.

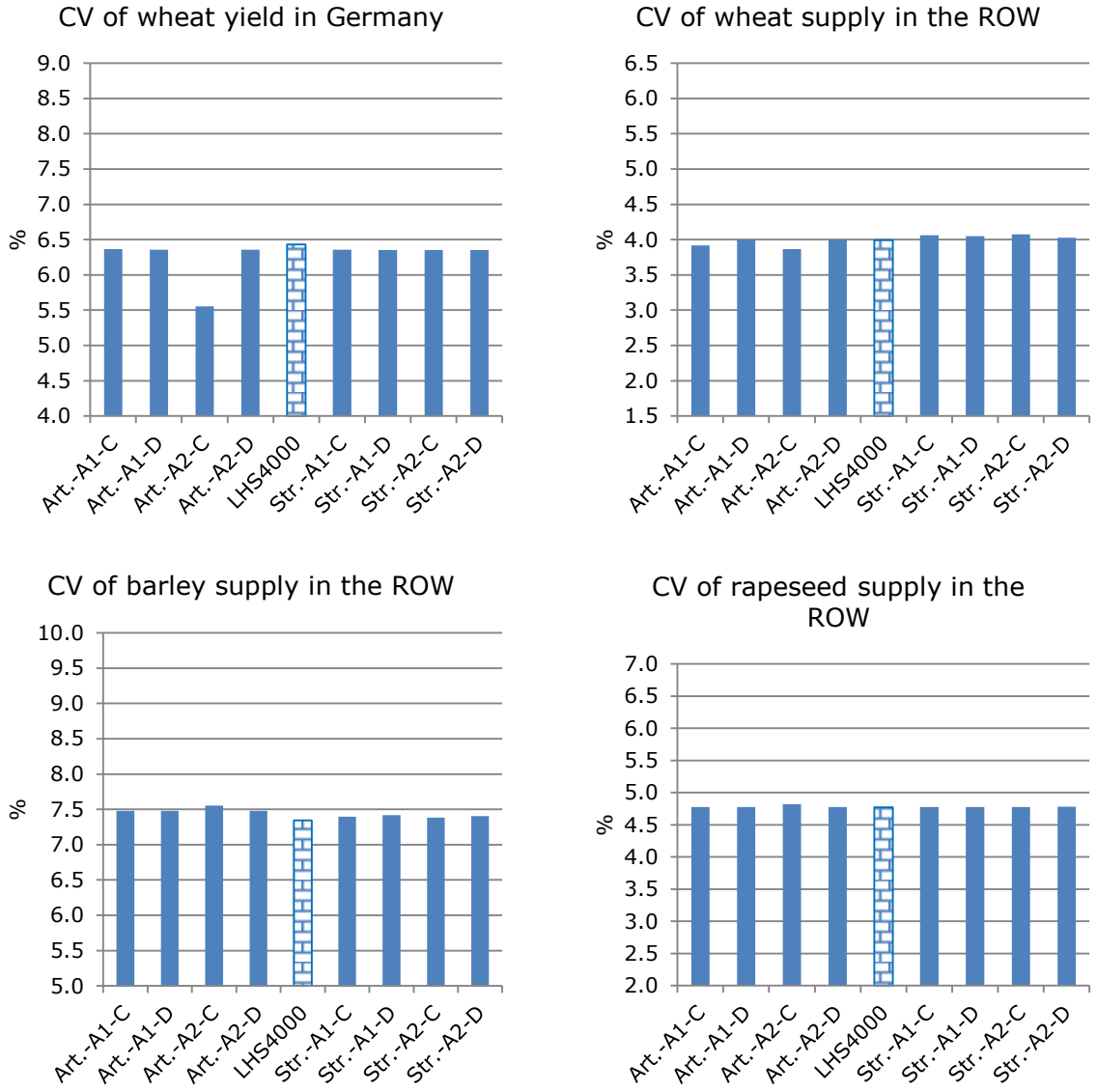
Figure 5.8 Expected value (EV) of wheat yield in Germany, wheat supply in the ROW, wheat prices in the EU and wheat prices in the ROW obtained with the different quadratures tested



Note: The names given to the quadratures are composed of the following elements: (Art.) denotes the n-octahedron from Stroud's theorem obtained with the formula from Artavia et al.; (Str.) denotes the n-octahedron from Stroud's theorem obtained with the formula from Arndt; (A1) and (A2) denote the arrangements 1 and 2 of the stochastic variables in the covariance matrix; (C) and (D) denote the methods to induce the covariance matrix to the quadratures, via Cholesky decomposition (C) and via the diagonalization method (D); LHS4000 indicates the sample obtained with Latin Hypercube Sampling of size $N = 4000$ and is the approximated 'true value'.

Source: own calculations

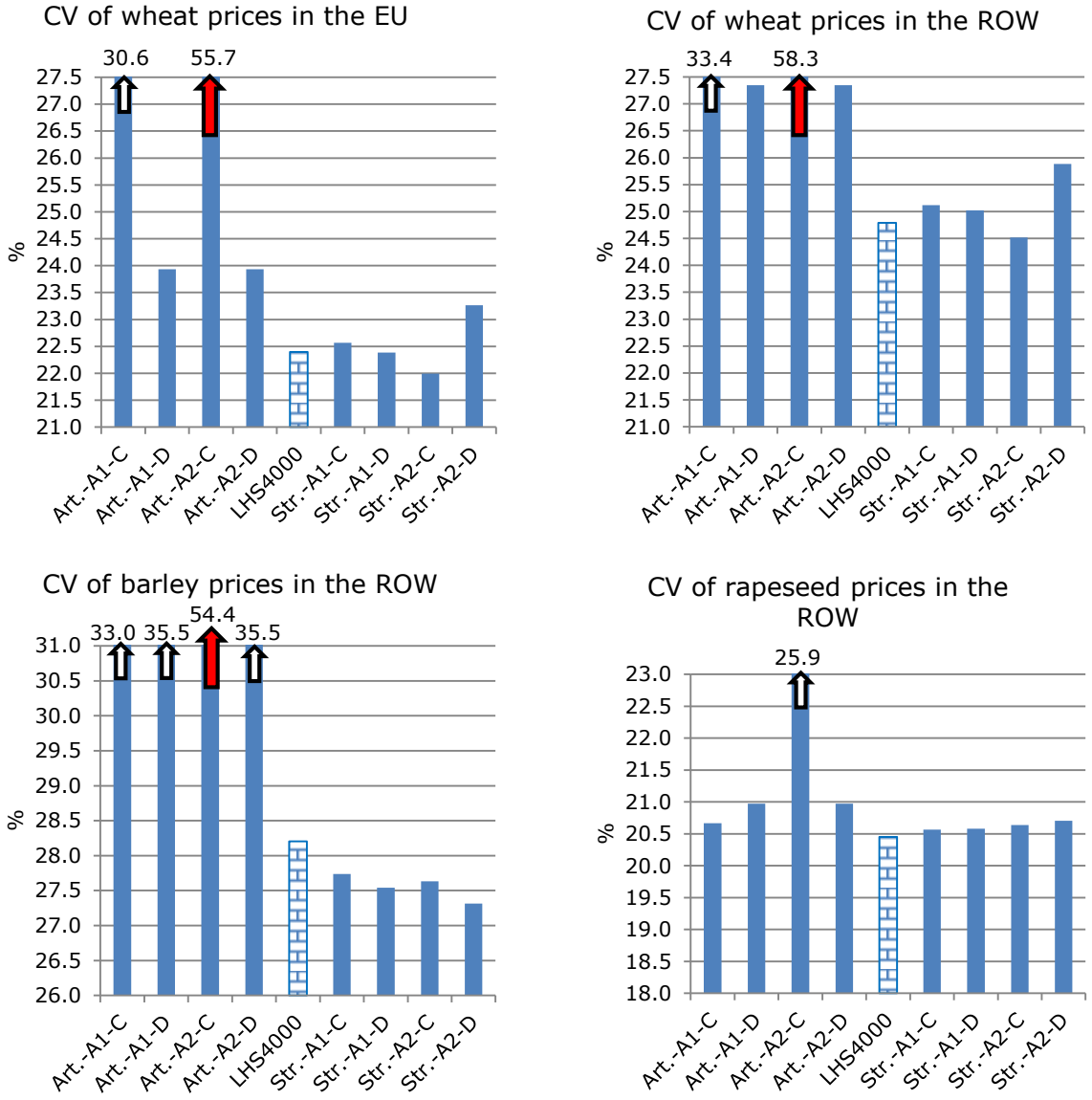
Figure 5.9 Coefficient of variation (CV) of wheat yield in Germany and of the supply of wheat, barley and rapeseed in the ROW obtained with the different quadratures tested



Note: The names given to the quadratures are composed of the following elements: (Art.) denotes the n-octahedron from Stroud's theorem obtained with the formula from Artavia et al.; (Str.) denotes the n-octahedron from Stroud's theorem obtained with the formula from Arndt; (A1) and (A2) denote the arrangements 1 and 2 of the stochastic variables in the covariance matrix; (C) and (D) denote the methods to induce the covariance matrix to the quadratures, via Cholesky decomposition (C) and via the diagonalization method (D); LHS4000 indicates the sample obtained with Latin Hypercube Sampling of size $N = 4000$ and is the approximated 'true value'.

Source: own calculations

Figure 5.10 Coefficient of variation (CV) of wheat prices in the EU and of the prices for wheat, barley and rapeseed in the ROW obtained with the different quadratures tested



Note: The names given to the quadratures are composed of the following elements: (Art.) denotes the n-octahedron from Stroud's theorem obtained with the formula from Artavia et al.; (Str.) denotes the n-octahedron from Stroud's theorem obtained with the formula from Arndt; (A1) and (A2) denote the arrangements 1 and 2 of the stochastic variables in the covariance matrix; (C) and (D) denote the methods to induce the covariance matrix to the quadratures, via Cholesky decomposition (C) and via the diagonalization method (D); LHS4000 indicates the sample obtained with Latin Hypercube Sampling of size $N = 4000$ and is the approximated 'true value'.

Source: own calculations

In Figures 5.8 to 5.10, it can be clearly seen that the different quadrature formulas result in different levels of accuracy of approximation of the model results. In Figures 5.8 and 5.9, it can be detected that, in general, the EV and CV of wheat yield in Germany and the supply of wheat, barley and rapeseed in the ROW are approximated well. In Figure 5.8, it can be observed that the EV of wheat prices in the EU and in the ROW is approximated accurately in most of the cases. Conversely, in Figure 5.10, we find that the CV of wheat prices in the EU and in the ROW, as well as the CV of prices for barley and rapeseed in the ROW, are approximated inaccurately by some of the quadrature formulas. For rapeseed, only the quadratures Art.-A2-C could not approximate well the results. The approximation of the CV of wheat and barley seems to be more difficult.

Core conclusions are:

i) the quadratures based on Arndt's formula (Str.) are more accurate than those based on Artavia et al.'s formula (Art.);

ii) for quadratures based on Arndt's formula and for wheat and barley, the difference between the arrangements of the coordinate system or the selected method to introduce correlation are important (for barley to a lesser extent);

iii) for quadratures based on Arndt's formula and for rapeseed, the differences between the arrangements or the methods to induce correlation are ambiguous; and

iv) for quadratures based on Arndt's formula and for barley, systematic lower coefficients of variations than the approximated 'true value' are obtained.

Several factors are identified which contribute to the explanation of the observed differences in accuracy:

i) the rotations tested result in very different samples of the stochastic space (put in other words, the quadrature formulas result in different discrete approximations of the marginal probability distributions of the stochastic variables included as factors to the supply and yield equations, see equations in Section 4.2);

ii) in ESIM, supply changes in large producing countries have stronger effects on prices than changes in small producing countries (asymmetrical consequences on prices of supply shocks in different countries); and

iii) in ESIM, the simulated supply shocks to the left and to the right in one country result in asymmetrical positive and negative price changes due to the functional forms of supply and demand used.

5.3.2 Analysis of Results

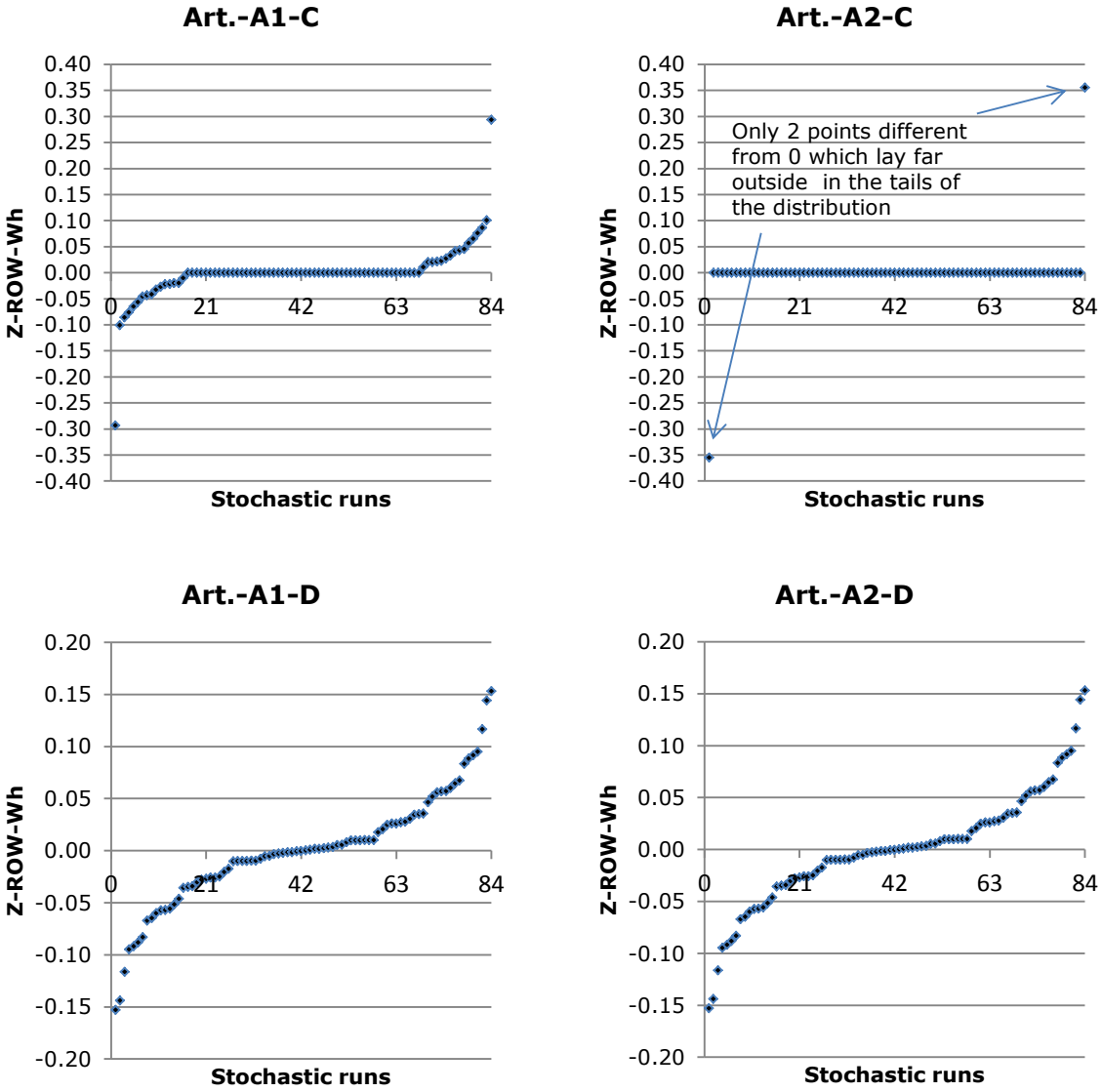
The analysis of the results is structured as follows: first, the factors which contributed to the differences in accuracy are exposed; second, the consequences of the factors on the determination of model results are discussed; and finally, the implications for the stochastic modeling with ESIM are elaborated.

Different Discrete Approximations of the PDFs of the Stochastic Variables

One reason for the differences in accuracy obtained is that the different rotations of the Q_t^n yield different discrete approximations of the PDFs of the stochastic variables; also, running ESIM over different quadrature points gives different results.

Figures 5.11 and 5.12 present the discrete approximations of the stochastic variable for wheat in the ROW obtained with the different rotations of the Q_t^n. This variable is presented since it is the largest country in ESIM and a supply shock to this variable results in significant price changes. The approximations are ordered from their lowest value to their highest (not in the order of appearance when running the model stochastically). For Art.-A1-C and Art.-A2-C, a broader scale is used for the vertical axes, so as to be able to depict the extremes of the disperse values generated. The discrete approximation obtained with the LHS4000 (the one used for the generation of the 'true value') is also shown in Figure 5.13 and serves as the reference point.

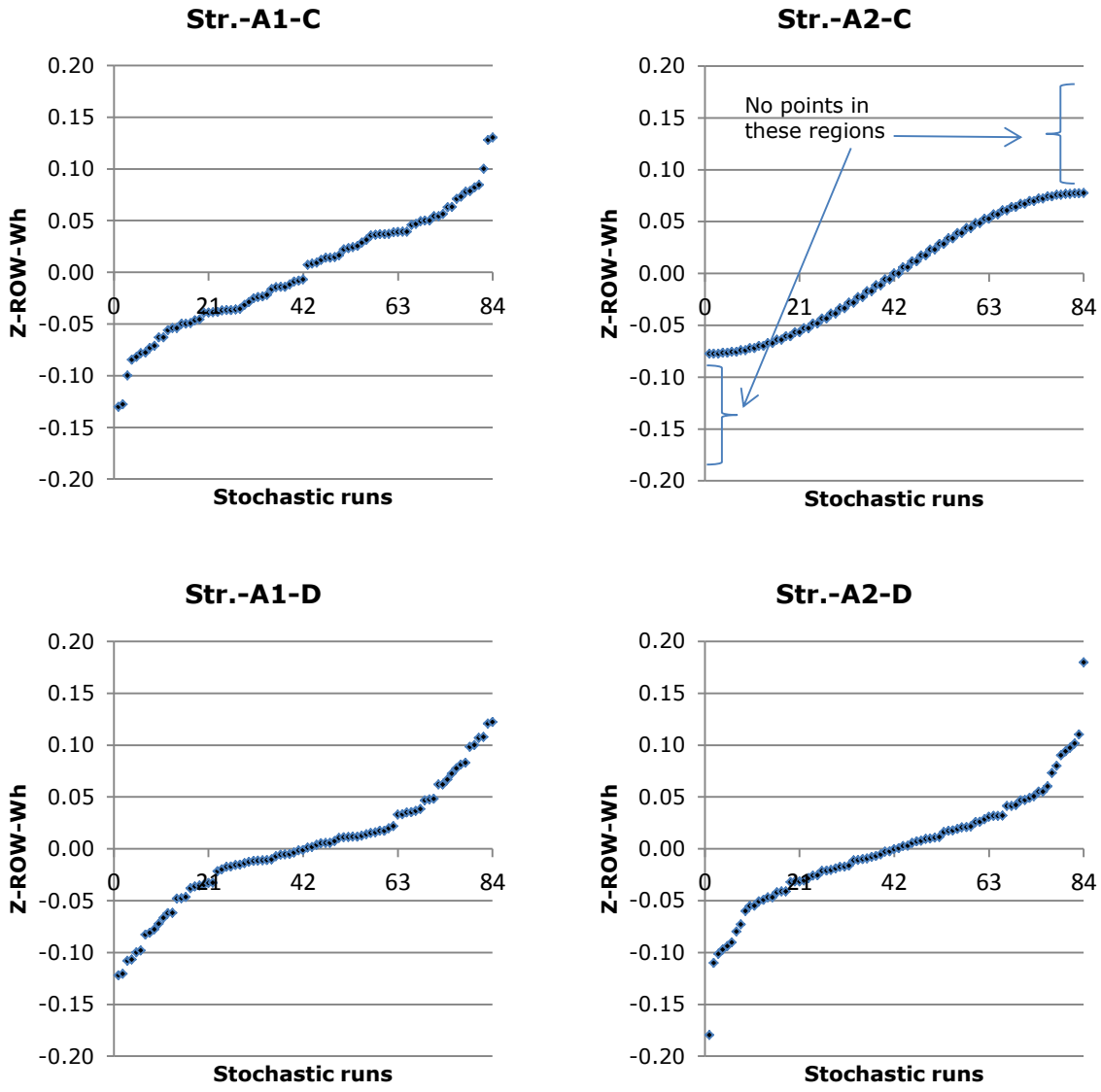
Figure 5.11 The discrete approximation of the PDF of the stochastic variable (in absolute terms) for wheat in the ROW obtained with Art.-A1-C, Art.- A2-C, Art.-A1-D and Art.-A2-D



Note: The names given to the quadratures are composed of the following elements: (Str.) denotes the n-octahedron from Stroud's theorem obtained with the formula from Arndt.; (A1) and (A2) denote the arrangements 1 and 2 of the stochastic variables in the covariance matrix; (C) and (D) denote the methods to induce the covariance matrix to the quadratures, via Cholesky decomposition (C) and via the diagonalization method (D). The stochastic variable is labeled with 'Z-ROW-Wh' and the units represent the percentage shock applied to the yield and supply equations in ESIM, e.g. 0.05 results in an increase of 5% of yield or supply.

Source: own development

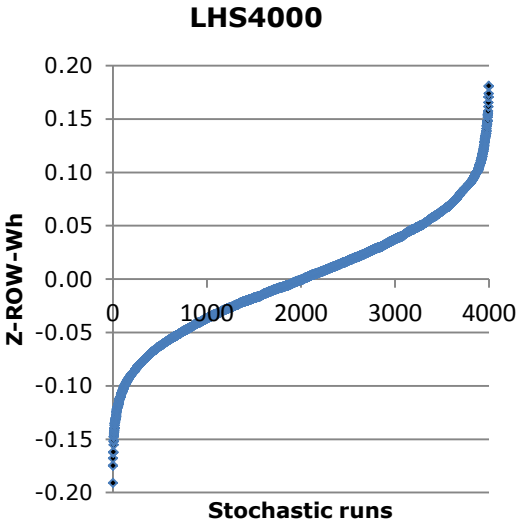
Figure 5.12 The discrete approximation of the PDF of the stochastic variable for wheat in the ROW obtained with Str.-A1-C, Str.- A2-C, Str.- A1-D and Str.- A2-D



Note: The names given to the quadratures are composed of the following elements: (Art.) denotes the n-octahedron from Stroud's theorem obtained with the formula from Artavia et al.; (A1) and (A2) denote the arrangements 1 and 2 of the stochastic variables in the covariance matrix; (C) and (D) denote the methods to induce the covariance matrix to the quadratures, via Cholesky decomposition (C) and via the diagonalization method (D). The stochastic variable is labeled with 'Z-ROW-Wh' and the units represent the percentage shock applied to the yield and supply equations in ESIM, e.g. 0.05 results in an increase of 5% of yield or supply.

Source: own development

Figure 5.13 The discrete approximation of the PDF of the stochastic variable for wheat in the ROW obtained with LHS4000



Note: LHS4000 indicates the sample obtained with Latin Hypercube Sampling of size $N = 4000$ and is the approximated 'true value'. The stochastic variable is labeled with 'Z-ROW-Wh' and the units represent the percentage shock applied to the yield and supply equations in ESIM, e.g. 0.05 results in an increase of 5% of yield or supply.

Source: own calculations

In Figures 5.11 and 5.12, it can be seen how the discrete approximations of the PDF of the stochastic variable for wheat in the ROW, obtained with the different rotations of the Q_t^n, differ strongly.

Figures 5.11 and 5.12 also show that, apparently, the rotation of the Q^n is important for the determination of the final rotation in the Q_t^n. It can be seen that the discrete approximations obtained using the Q^n from Artavia et al. are more flat in the middle and more disperse on the ends of the graphs. This can be easily detected in the upper two graphs of Figure 5.11. However, this also occurs for the lower graphs of that figure, but less pronounced. This can be seen in Runs 42 to 63 when comparing them to LHS4000 in Figure 5.13. On the other hand, the discrete approximations obtained with the quadrature formulas from Arndt (Str.) are steeper in the middle and less disperse on the end of the graphs. It seems that the transformation of the Q^n does not result in such a big rotation; that is, that the transformation induces the desired covariance matrix, but, on the whole, the vertices of the octahedron remain close to their original position. This can be seen by analyzing the graphs in Figure 5.12. The quadratures Str.-A2-C can be used as reference for the comparison since the sample of Z-ROW-Wh is the same before and after the transformation but is contracted (or extended) to induce the variance of the stochastic variable. It can be seen that the samples of Z-ROW-Wh obtained with Str.-A1-C, Str.-A1-D, and Str.-A2-D are all similar

Str.-A2-C. This leads us to think that with the transformation, the Q^n is adapted to fit the new stochastic space but it is not rotated much. The new stochastic space is an n-ellipsoid which considers the covariance between the variables and is no longer the n-sphere for independent variables. If the stochastic variables all presented equal variance, then with low correlations the n-ellipsoid would be similar to an n-sphere; with a strong correlation, the n-ellipsoid would simulate more the form of an American football ball.

As mentioned above, the sample of Z-ROW-Wh obtained with Str.-A2-C is the same before and after the transformation applied to the Q^n. This also happens with the quadrature Art.-A2-C. This occurs since with the transformation via Cholesky decomposition, the first variable in $\mathbf{Z}$ is only contracted or expanded to simulate the variance of the stochastic variable, but it is not rotated. This is a result of the multiplication of the lower triangular matrix $\mathbf{L}$ with the matrix of standard normal quadratures $\mathbf{\Gamma}$, which occurs since in the first row in $\mathbf{L}$ only the first value is different from zero (see Box 3.8).

As a result, with Art.-A2-C, the first variable in $\mathbf{Z}$ is approximated with $\pm\sigma_i\sqrt{n}$ and the rest of points are zeros. Since $n = 42$, then $\pm\sigma_i\sqrt{n} \approx \pm\sigma 6.48$, which are points lying far outside in the tails of the distribution (see upper right graph in Figure 5.11). If we consider the interval of integration for normal distributed variables as $[-3\sigma, 3\sigma]$, then, with it we would cover about 99.7% of the distribution (Gujarati, 2003). With this definition of the region of integration, the points $\pm 6.48\sigma$ would be lying outside the region of integration. These points are very unlikely in reality and their application in ESIM results in longer time requirements to find solutions, since the system of equations is forced to very extreme situations. For example, with the rotations obtained with Arndt's formula, ESIM requires about 35 min. to run; with the rotations obtained with Artavia et al.'s formula, ESIM requires about 60 min. to run.

With Str.-A2-C, the stochastic variable is sampled with points which are inside the interval $\left[-\sigma_i\sqrt{2}, \sigma_i\sqrt{2}\right]$ and which avoid the tails of the normal distribution (see upper right graph in Figure 5.12). Consequently, with the linear transformation via the Cholesky decomposition, the first stochastic variable in $\mathbf{Z}$ is systematically approximated with points which do not correspond to the normal distribution very well.

By analyzing the samples of all the stochastic variables, we find that for the quadratures using Artavia et al.'s formula, 48% of the variables are approximated with extreme values that are at least 20% lower and 20% higher than the reference min and max values of LHS4000, respectively. For the quadratures using Arndt's formula, 98% of the discrete approximations present min and max points that are at least 20% higher and 20% lower than the reference min

and max values respectively; and thus, interior to the interval covered by the LHS4000 sample.

By comparing the graphs from the left side with the graphs from the right side in Figures 5.11 and 5.12, it can be detected that the linear transformations using the Cholesky decomposition and the diagonalization method are coordinate dependent. The discrete approximations obtained with the arrangement A1 differ from those obtained with the arrangement A2. In Figure 5.11, it can also be noted that Art.-A1-D and Art.-A2-D are equal. This occurs because with the diagonalization method, the eigenvectors and the eigenvalues of $\boldsymbol{\Sigma}[\boldsymbol{z}]$ are computed and these do not change with the different arrangements of the coordinates. Nonetheless, the coordinates of the eigenvectors do change. Since in the Q^n obtained with Artavia et al.'s formula, all points lie on the coordinate axes, the change of the coordinates of the eigenvectors does not have an effect on the final rotation of the Q_t^n; thus, Art.-A1-D and Art.-A2-D are equal. Conversely, the Q^n obtained with the quadrature formula from Arndt is rotated; thus, the change of the coordinates of the eigenvectors results in different rotations of the Q_t^n.

Asymmetries in ESIM

From Chapter 2, we know that the accuracy with which the numerical integration formulas approximate the integral $\int_{\mathbb{R}^n} g(\mathbf{x})f(\mathbf{x})\,d\mathbf{x}$ depends not only on the degree of precision of the formula but also on the accuracy with which $g(\mathbf{x})$ itself can be approximated by polynomials (Haber, 1970). In this sense, the functions in ESIM appear to be complex and not easily approximated by polynomials. Some model structures have been identified which may be influencing the accuracy of approximation of the model results with 'efficient quadratures'; these are:

i) the existence of large regions (e.g. the ROW), whose changes strongly determine the outcome of the model (e.g. prices);

ii) the isoelastic supply and demand curves which are very inelastic; and

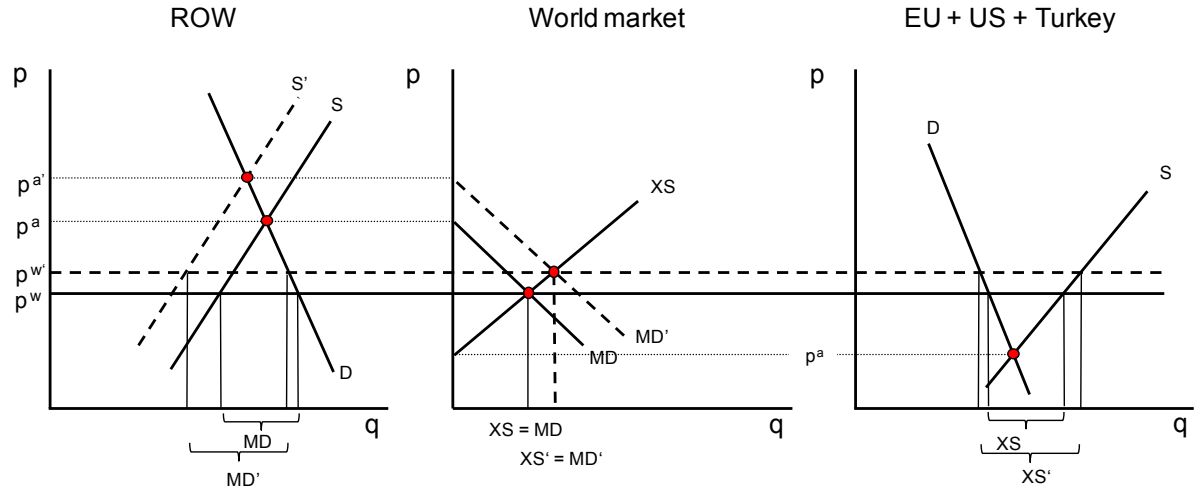
iii) the threshold points set up by a protectionist price policy system for some commodities.

These model structures are explained below.

i) The existence of large regions (e.g. the ROW), whose changes strongly determine the outcome of the model (e.g. prices).

In order to understand the price mechanism and how large the ROW is in ESIM, see Figure 5.14 and Table 5.2.

Figure 5.14: A simplification of the price determination mechanism in ESIM



Note: The notation used has the following meaning: $p^{a:}$ price with autarky; p^{w}: world market price (domestic price under free trade); S: supply curve; D: demand curve; XS: total export supply curve; MD: total import demand; ‘: situation after shock

Source: own development

The above figure shows how, in ESIM, prices are obtained at the equilibrium points where the world net export supply equals the world net import demand. Note, that price changes are produced by shifts of the export supply or import demand curves in the world market, and note that these shifts are the result of either shifts of the supply or demand curves of large countries or of total shifts obtained from the sum of all shifts at the country level. Shocks on the supply curves of countries with low production, in the world market context, represent small shifts of the net export supply curve and the net import demand curve. For example, a shift of 5% in the supply of wheat in Belgium results in a very small shift of the supply curve for the EU + US + Turkey in Figure 5.14. On the other hand, a 5% shift of the supply curve of the ROW, e.g. to the left as in the figure, results in a significant effect on net import demand in the world market, and, thus, in a significant effect on world market prices. For this reason, the stochastic variables for the Rest of World play a primary role in the determination of prices in ESIM. As a result, it is crucial for the approximation of model results to generate discrete approximations of the PDFs of the stochastic variables for the ROW that not only match the central moments of the observed PDF but also correspond to the bell curve shape of the normal distribution. The effects of discrete approximations that do not correspond to the bell curve are analyzed below, in the discussion of results.

For the crops with stochastic variables, Table 5.2 summarizes the world’s biggest suppliers.

Table 5.2: Countries with high production levels of wheat, barley and rapeseed in ESIM in the year 2015 in the baseline scenario (in % of world market supply)

	Wheat	Barley	Rapeseed
EU-27 (> 3% of World Production)*	(%)	(%)	(%)
FR	5.1	7.4	7.0
GE	3.4	8.3	9.6
UK	2.2	3.9	3.5
PL	1.3	2.4	3.4
ES	0.6	6.5	-
World Market			
TU	2.9	6.5	-
US	9.0	3.7	1.4
EU	17.9	41.2	29.1
ROW	70.2	48.7	69.5

* Countries with more than 3% of world production in at least one of the listed crops

Note: The countries are abbreviated as follows: FR: France, GE: Germany, PL: Poland, ROW: 'rest of the world', ES: Spain, TU: Turkey, UK: United Kingdom, US: United States of America.

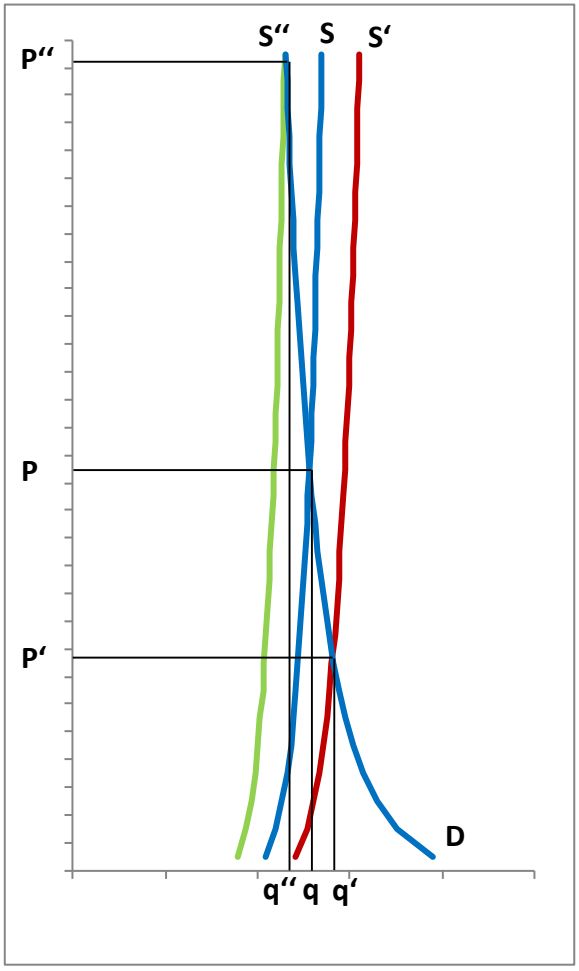
Source: own calculations

Table 5.2 shows that the ROW is clearly the biggest producer of wheat, barley and rapeseed among the 30 countries or country aggregates depicted in ESIM.

ii) The isoelastic supply curves which are very inelastic.

The second important asymmetry in ESIM is that with isoelastic supply and demand functions, positive and negative shocks of equal magnitude result in price changes of different sizes. These differences are particularly strong with inelastic curves. This is shown in Figure 5.15.

Figure 5.15 Example of the price effect of inelastic supply and demand functions



Note: (P) denotes the price at the equilibrium point; (S) denotes the supply curve; (D) denotes the demand curve; and the single and double apostrophe (‘) and (‘‘) denote the situations after rightward and leftward shocks of equal magnitude.

Source: own development

Such a situation as that depicted in Figure 5.15 occurs especially with the market for barley in ESIM. The reaction of demand in the ROW, a region that covers around 70% of world demand, is very low. This explains why for all quadratures obtained with Arndt's formula, the approximated CVs of barley prices are below the approximated true value (see bottom left graph in Figure 5.10). The samples obtained with those quadratures avoid points on the tails of the distributions that would result in large price changes. Without those points, the generated CVs are smaller.

iii) The threshold points set up by the protectionist price policy system for some commodities.

The final model structure identified that may be influencing the accuracy of the quadratures is the system of price policies simulated. These result in threshold

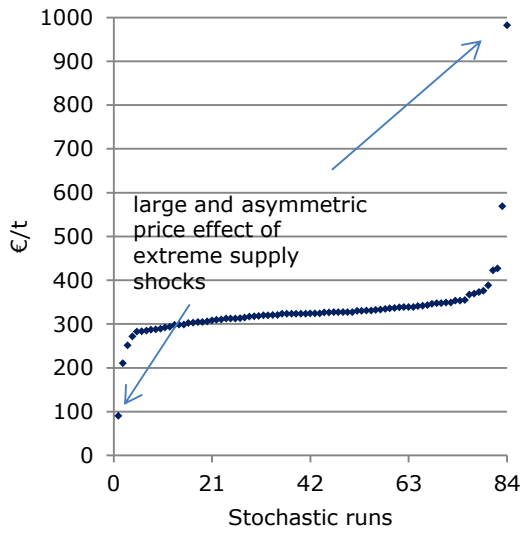
points where price levels change abruptly (see Figures 4.1 - 4.4), making the approximation of model results more difficult. With samples of the stochastic space being of small size, as is the case with the quadratures tested, it may be that points which would have triggered the price policies remain unevaluated or, alternatively, that only points which trigger the polices are evaluated. These situations would result in inaccuracies. The effect of the price policies simulated for wheat on the accuracy of the results obtained with LHS samples of increasing size is tested in Section 5.1. It is clearly possible to observe that the policies make the numerical integration more complex.

Discussion of Results

A selection of the results from Figures 5.8 to 5.10 are discussed below, taking into consideration: a) the different discrete approximations of the PDFs of the stochastic variables obtained with the different quadrature formulas (the different rotations of the Q_t^n), and b) the asymmetries identified in ESIM. Together, these factors help explain the differences in the accuracy of approximation of the model results of the different rotations tested.

The quadrature formula Art.-A2-C is the only one not able to approximate well the CV of rapeseed prices. The discrete approximation of the PDF of the stochastic variable for rapeseed in the ROW generated with Art.-A2-C is very similar to the discrete approximation of the PDF of the stochastic variable of wheat in the ROW (see Figure 5.11). The difference is that in this case the extreme values are ± 0.45, which corresponds to $\pm\sigma_i\sqrt{n}$. The extreme supply curve shifts for the ROW result in very large and asymmetric price effects, as shown in Figure 5.15. The stochastic runs in ESIM with Art.-A2-C produced the following rapeseed prices presented in Figure 5.16.

Figure 5.16 Rapeseed prices in ESIM with the quadratures Art.-A2-C



Source: own development

The very disperse price values obtained for Runs 1, 2, 83 and 84 (65% to 300% of the EV) in Figure 5.16 explain the higher CV of rapeseed prices observed for Art.-A2-C. The other quadrature formulas better approximate the CV of rapeseed prices, because their discrete approximation of the PDF of the stochastic variable for rapeseed in the ROW does not include the very disperse points from the mean. Their generated discrete points correspond more with the normal distribution from the PDF of the stochastic variable.

Note that in general, the quadrature Art.-A2-C is the less accurate (see Figures 5.8 to 5.10). The reason is that the arrangement A2 has the variables of the ROW (wheat, barley and rapeseed) in the first three positions, and that the transformation of Artavia et al.'s quadratures via the Cholesky decomposition result in samples of the first's stochastic variables with a few extreme values lying in the latter ends of the tails of the marginal distributions and a concentration of points close to the mean (similar to the sample of wheat in the ROW obtained with Art.-A2-C, as shown in Figure 5.11).

Another result which can be observed in Figure 5.10 is that it is more difficult to approximate the CV of the prices of wheat and barley than the CV of rapeseed prices. This result was already observed and discussed in Section 5.1. There, the main identified factor is the system of price policies for wheat. Also, it is exposed that another factor may be the stronger cross relationships between wheat and barley, which result in an increase of the stochastic dimensions with an influence on the prices of those commodities. Furthermore, Table 5.2 shows that wheat is particularly asymmetrical for a large region (ROW), a medium size country (US), and some other rather small countries in the EU. Thus, the

approximation of the probability distribution of the stochastic variables of the large regions is of special importance, since changes in supply in these regions result in significant changes in prices. This characteristic of the wheat market contributes to the explanation of the difference in accuracy obtained with the quadratures based on Arndt's formula and arrangements A1 and A2 of the covariance matrix **Z**. The samples of the probability distributions of the stochastic variables of these large regions appear to be very significant in the determination of prices, and, thus, the small changes in the allocation of the points of the samples originated by A1 and A2 may result in different coefficient of variations of wheat prices (see the graph for wheat prices in the ROW in Figure 5.10). Note also, that significant differences are observed for the quadratures Str.-A2-C and Str.-A2-D. This indicates that, again, if some regions strongly influence the developments of one market in the model, then small changes in the stochastic sample used for it may result in different approximations of the moments of model results (e.g. the coefficient of variation of prices).

In ESIM, barley is also asymmetrical, but from the demand side; as mentioned above, the ROW covers around 70% of world demand and the demand is very inelastic. This results in strong price reactions to small shocks and makes the approximation of model results more difficult.

Accuracy of the approximation of the CV of rapeseed prices is easier to achieve, since this market has fewer asymmetries. Unlike the case for barley and wheat, neither world supply nor world demand are heavily concentrated in only one region of the model. Furthermore, the demand functions are the most elastic, which results in more symmetrical price changes.

In Figures 5.9 and 5.10, it can also be seen that, in general, the quadratures generated in combination with Artavia et al.'s formula perform poorly, while the quadratures using Arndt's formula (Str.) are close to the estimated true value. This occurs because the quadratures using Artavia et al.'s formula generate discrete approximations of the PDFs of stochastic variables that are more flat in the middle (near the mean) and more disperse in the tails (referring to the form when the discrete approximations are ordered from their lowest value to their highest, as in Figures 5.11 – 5.13). These discrete approximations of the PDFs of stochastic variables in combination with the effect of inelastic isoelastic curves shown in Figure 5.15 result in the higher EV and CV of the model results observed in Figures 5.8 and 5.10. Furthermore, if the stochastic variable sampled with those characteristics shock the supply variables of large countries in ESIM (such as ROW), then the approximation errors are transmitted to the results of other variables through the rich cross relationships and through the world market price.

The quadratures based on Arndt's formula are more accurate since the samples obtained with it present neither a concentration of points close to the mean or any other point nor the extreme values at the latter ends of the tails of the marginal distributions.

Implications for the Stochastic Modeling with ESIM

The results above show that for models with strong asymmetries like wheat in ESIM, different quadrature formulas result in different approximations of the model results. With ESIM and in general, it is desirable to choose a formula that generates discrete approximations of the PDFs of the stochastic variables that not only match the moments but also correspond to the observed normal PDF. Put another way, that the points are inside the interval $[-3\sigma, 3\sigma]$, which covers about 99.7% of the distribution, and if possible that each point represents an interval of equal probability of the input distribution, as is done in the LHS method.

Out of the 8 possible quadrature formulas, it is shown that the rotation of the Q^n obtained with Arndt's formula (Str.) performs better than when the vertices are lying on the coordinate axes (Artavia et al.'s proposition). For the quadratures from Arndt, it is demonstrated that for markets with asymmetries, the differences in the samples produced by the arrangements A1 and A2 or by the methods to induce the covariance matrix via the Cholesky decomposition (C) or the diagonalization method (D) are important, but it is not possible to identify one as the best. However, it was explained that the linear transformations using the Cholesky decomposition systematically approximate the stochastic variables ordered first in $\mathbf{Z}$ in a way which does not correspond to normal distributions. Since with A2 the variables of the ROW that have a strong effect in the determination of prices in ESIM are ordered first in $\mathbf{Z}$, it was decided not to use the quadrature Str.-A2-C. From the remaining three quadratures, no distinction is identified; thus, it is considered that any of them can be used for further analysis.

6 The Consequences of the Liberalization of the EU Cereals Regime on Market Instability

In this chapter, the relevance of multi-market stochastic modeling is demonstrated by showing the consequences of the liberalization of the EU cereals regime on the instability of EU and world markets, as well as illustrating the information to be gained by considering uncertainty. The effect of the simulated price policies on the EU domestic prices with different EU net trade positions is also shown.

The focus of the analysis is the demonstration of the possibilities of stochastic commodity market modeling, and not the detailed depiction of the political scenarios.

For the analysis, the stochastic version of ESIM with the quadratures Str.-A1-D is used.

6.1 The Simulated Scenarios

The exercise utilizes two scenarios. The first one is the baseline (Bl) scenario for the projection horizon 2006-2015. The model is calibrated to reproduce world market price projections of large-scale models covering the developments of world markets like the FAPRI or the AGLINK model (Banse, 2005). The baseline scenario also reflects the agricultural and trade policies of the EU. However, this chapter specifically concentrates on the price policies applied to the markets of wheat, barley and rapeseed in the EU. The simulated price policies for those markets in the baseline scenario are shown in Table 6.1.

Table 6.1: The simulated price policies for the markets of wheat, barley and rapeseed in the EU in 2015 in the baseline (Bl) scenario

		Wheat	Barley	Rapeseed
Intervention price	€	88.9	-	-
Specific export subsidy	€/t	5.1	5.6	-
Export subsidy limit	Mio t	13.4	8.6	-
Tariff rate quota (TRQ)	Mio t	2.3	0.3	-
Threshold price	€	137.8	137.8	-

Note: The intervention price and the threshold price are deflated using 2007 as the base year. Thus the values differ from the 101.31 €/t (intervention price) and 157 €/t (threshold price) stated in the EU cereals regime.

Source: own development based on the ESIM version[8] used for this exercise

Table 6.1 shows that for the wheat market, the price policies simulated are: an intervention price, a specific export subsidy, an export subsidy limit, a TRQ, and a threshold price. For barley, the intervention price is not simulated, because since the start of the 2010/11 marketing year the buying-in ceiling has been set to zero (European Commission, 2011). Rapeseed is simulated as completely liberalized.

The second scenario is the free trade (Ft) scenario in which all price policies are abolished.

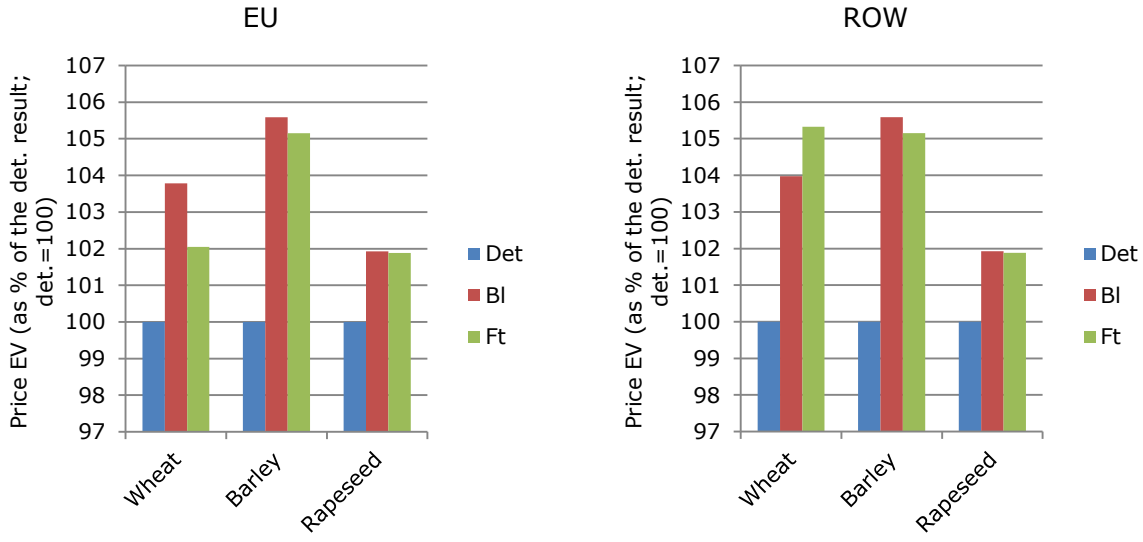
6.2 Results and Discussion

6.2.1 Presentation of Results

The stochastic version of ESIM was run simulating the baseline (Bl) and the free trade (Ft) scenarios. A selection of the results are presented in Figures 6.1 to 6.3; Figure 6.1 exhibits the obtained prices (EV) in relation to the deterministic results, Figure 6.2 focuses on the consequences of free trade on price levels using the baseline as the reference scenario, and Figure 6.3 illustrates the consequences of free trade on the uncertainty of prices. Note that the vertical axes of the graphs in Figures 6.2 and 6.3 are divided into units of 0.5%, and that for Figure 6.2, the 0.5% is calculated from the result of the baseline scenario for the ROW. This enables comparisons of the consequences of free trade between commodities.

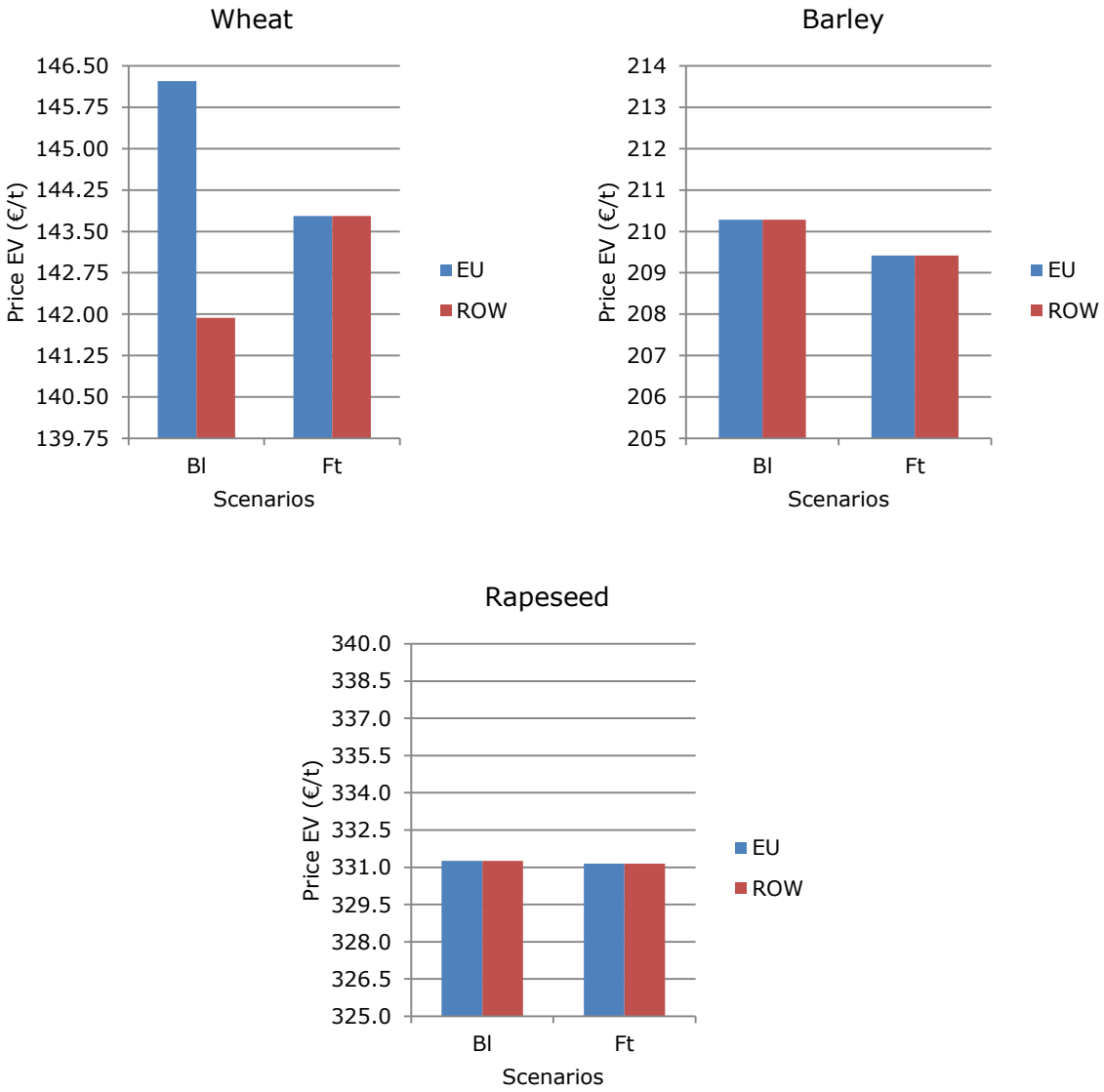
[8] ESIM is continuously developed; thus, the policies modeled may vary in time.

Figure 6.1 The expected value (EV) of EU and world market prices obtained with the baseline (Bl) and the free trade (Ft) scenarios as percentages of the deterministic (Det) result



Source: own calculations

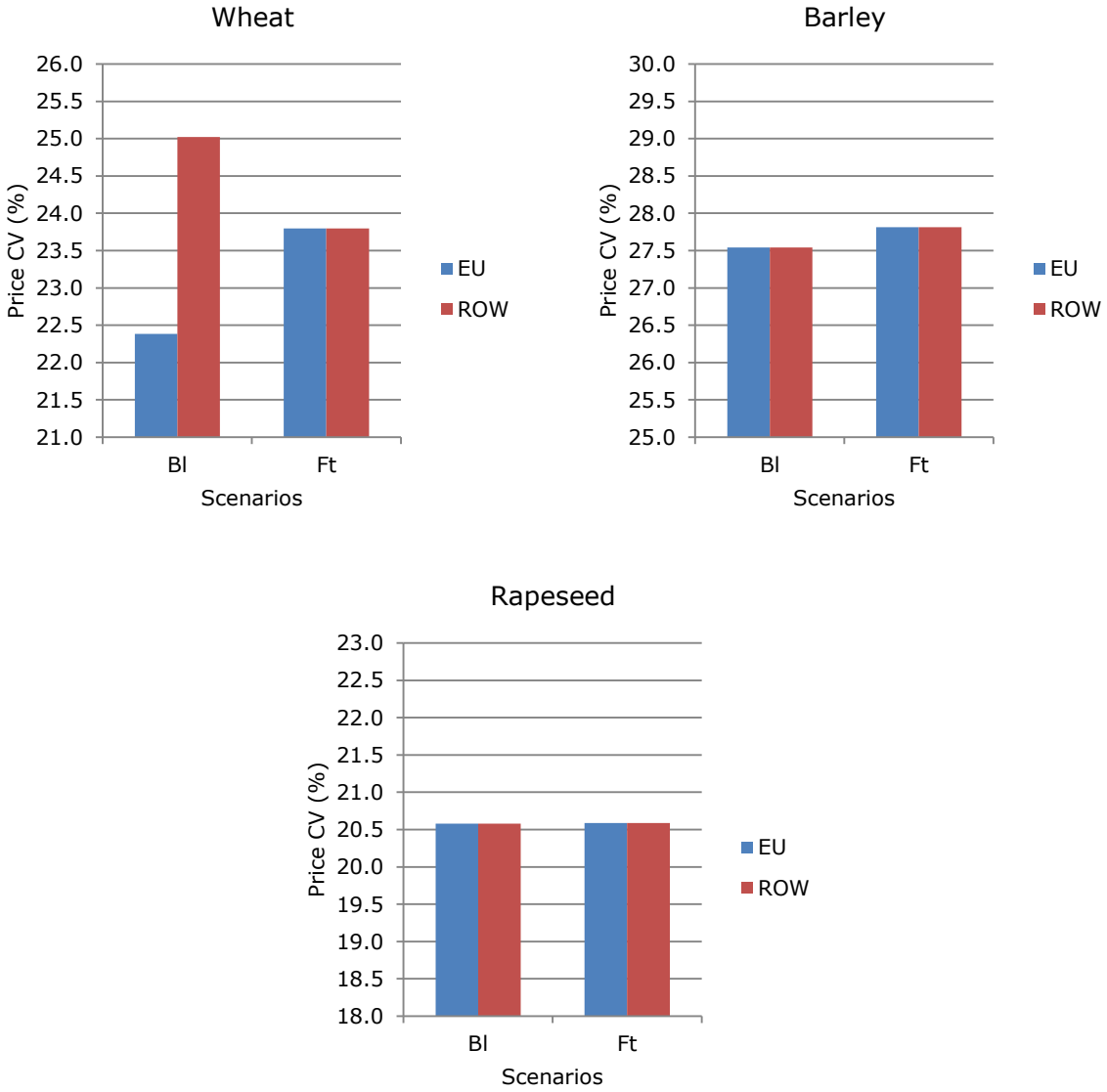
Figure 6.2 The expected value (EV) of wheat, barley, and rapeseed prices in the EU and the ROW obtained with the baseline (Bl) and the free trade (Ft) scenarios



Note: The vertical axes of the graphs are divided into units of close to 0.5% of the ROW results obtained with the baseline scenario.

Source: own calculations

Figure 6.3 The coefficient of variation (CV) of wheat, barley, and rapeseed prices in the EU and the ROW obtained with the baseline (Bl) and the free trade (Ft) scenarios



Source: own calculations

The core conclusions are:

i) The EVs of the prices obtained with the stochastic version of ESIM are higher than the obtained prices with the deterministic run of the model. The largest difference is observed for barley, followed by wheat, while rapeseed presents the smallest difference.

ii) Under yield uncertainty, the EU price policies are triggered only in the wheat market and these result in higher EU price levels (with respect to world market prices). Liberalization of the market results in lower wheat prices in the EU and higher in the ROW. The EU and the ROW markets for barley also react to the liberalization of wheat price policies with a small decrease on prices. Rapeseed price levels do not react at all.

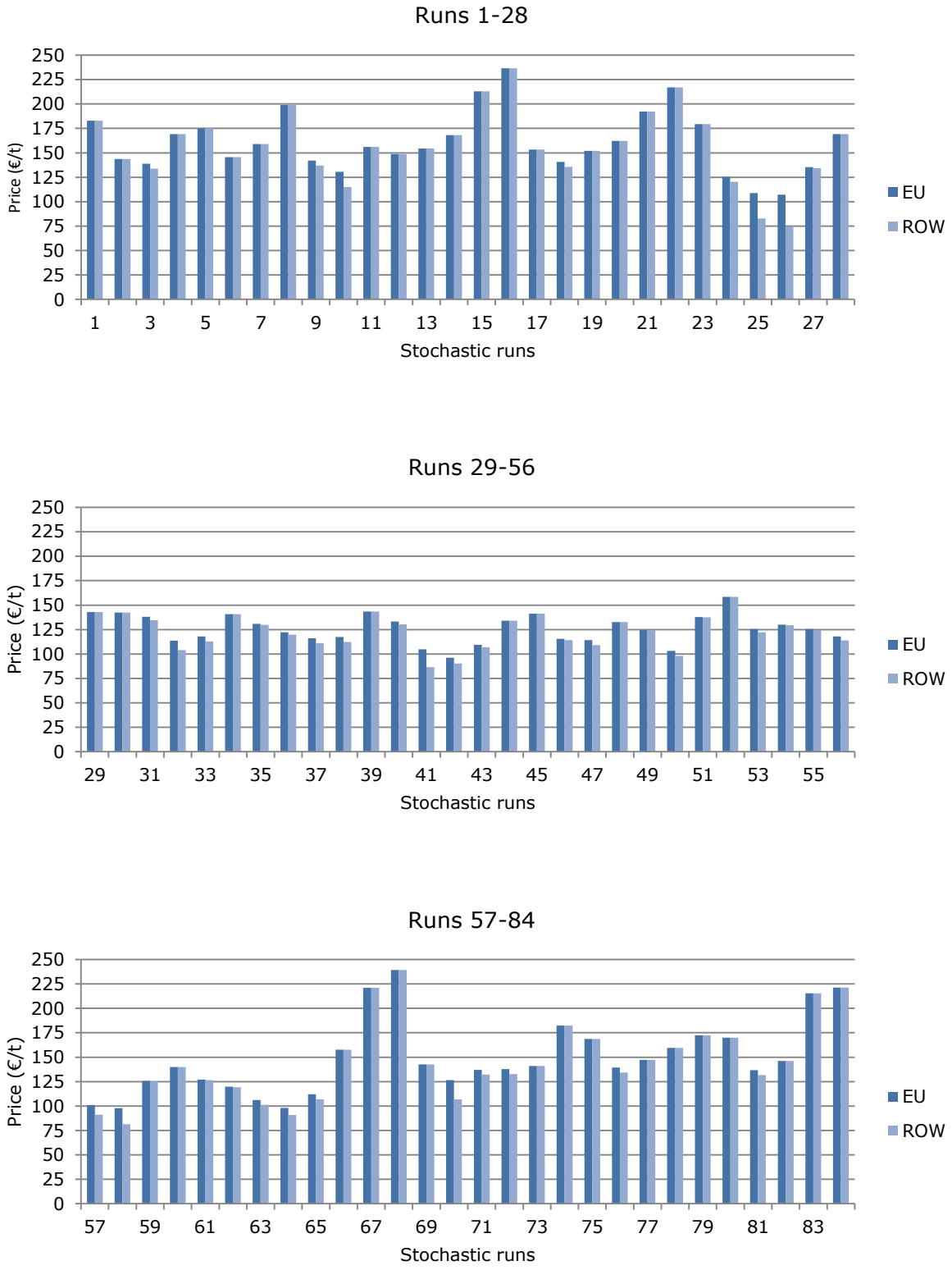
iii) Free trade results in an increase in wheat price uncertainty in the EU (CV of wheat prices increases with free trade) and a decrease in the ROW.

6.2.2 Analysis of Results

The EVs of the prices obtained with the stochastic version of ESIM are higher than the obtained prices with the deterministic run, due to the inelastic supply and demand curves as explained in Section 5.3.2 (see Figure 5.15). The effect is different between the commodities because the more inelastic the curves are, the stronger is the positive price effect. For barley, the demand elasticity in the ROW is very low, resulting in the strongest price effects. For rapeseed, the demand curves are the most elastic, resulting in the lowest positive price difference (with respect to the deterministic result).

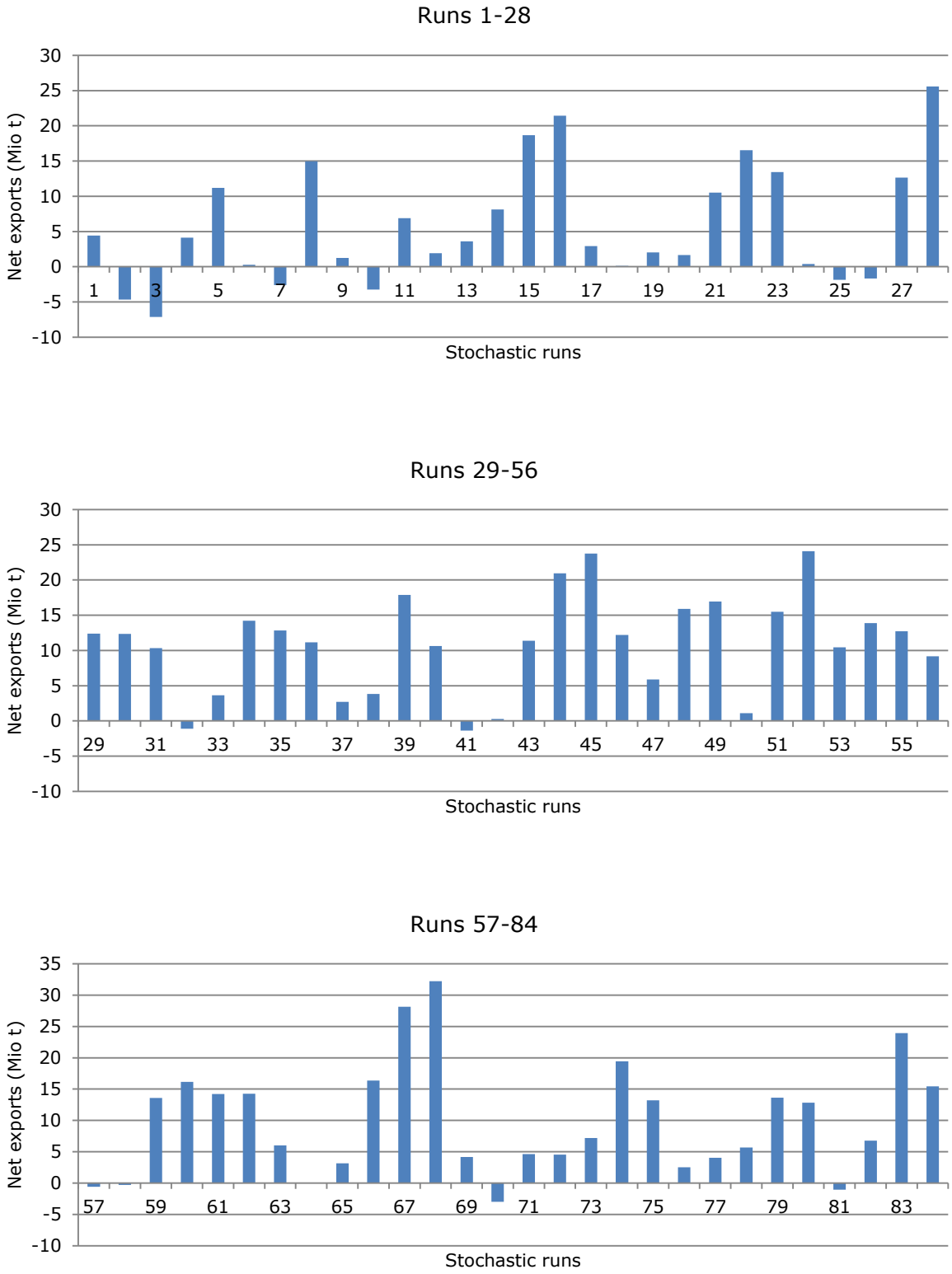
Figures 6.4 to 6.8 support the core conclusion (ii) from the above list. Figures 6.4 and 6.6 show for wheat and barley, the EU and the ROW prices obtained in each of the stochastic runs. Different price levels indicate the cases in which price policies are applied. Figures 6.5 and 6.7 present the EU wheat and barley net trade positions, which determine which price policy is applied, obtained in each run. An additional factor determining the application of the price policies is the level of the world market prices. The EU wheat market price system in the case of world market prices being below the intervention price is shown in Figure 6.8.

Figure 6.4 EU and ROW wheat prices obtained in each of the stochastic runs in the baseline scenario



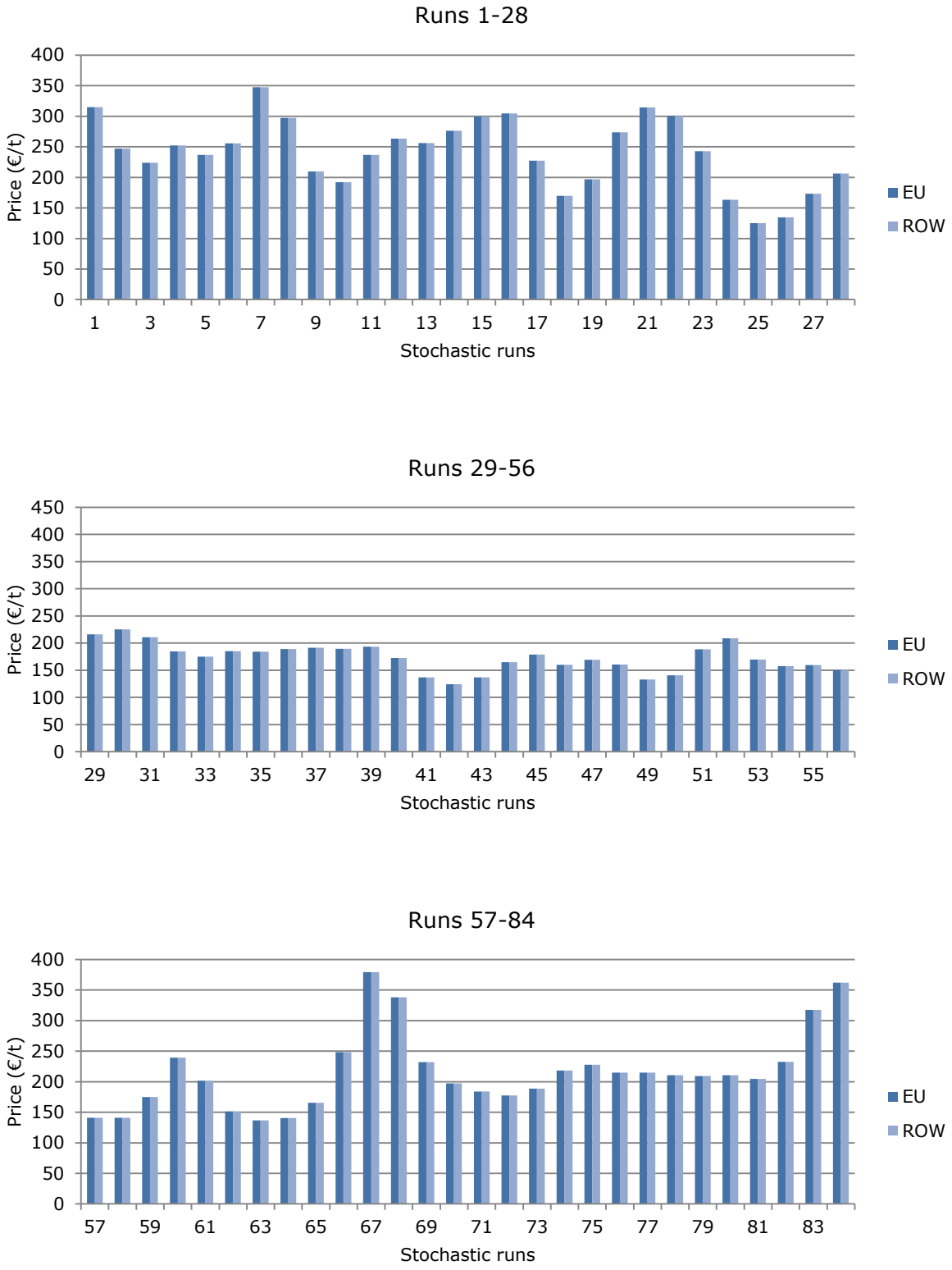
Source: own calculations

Figure 6.5 EU wheat net exports obtained in each of the stochastic runs in the baseline scenario



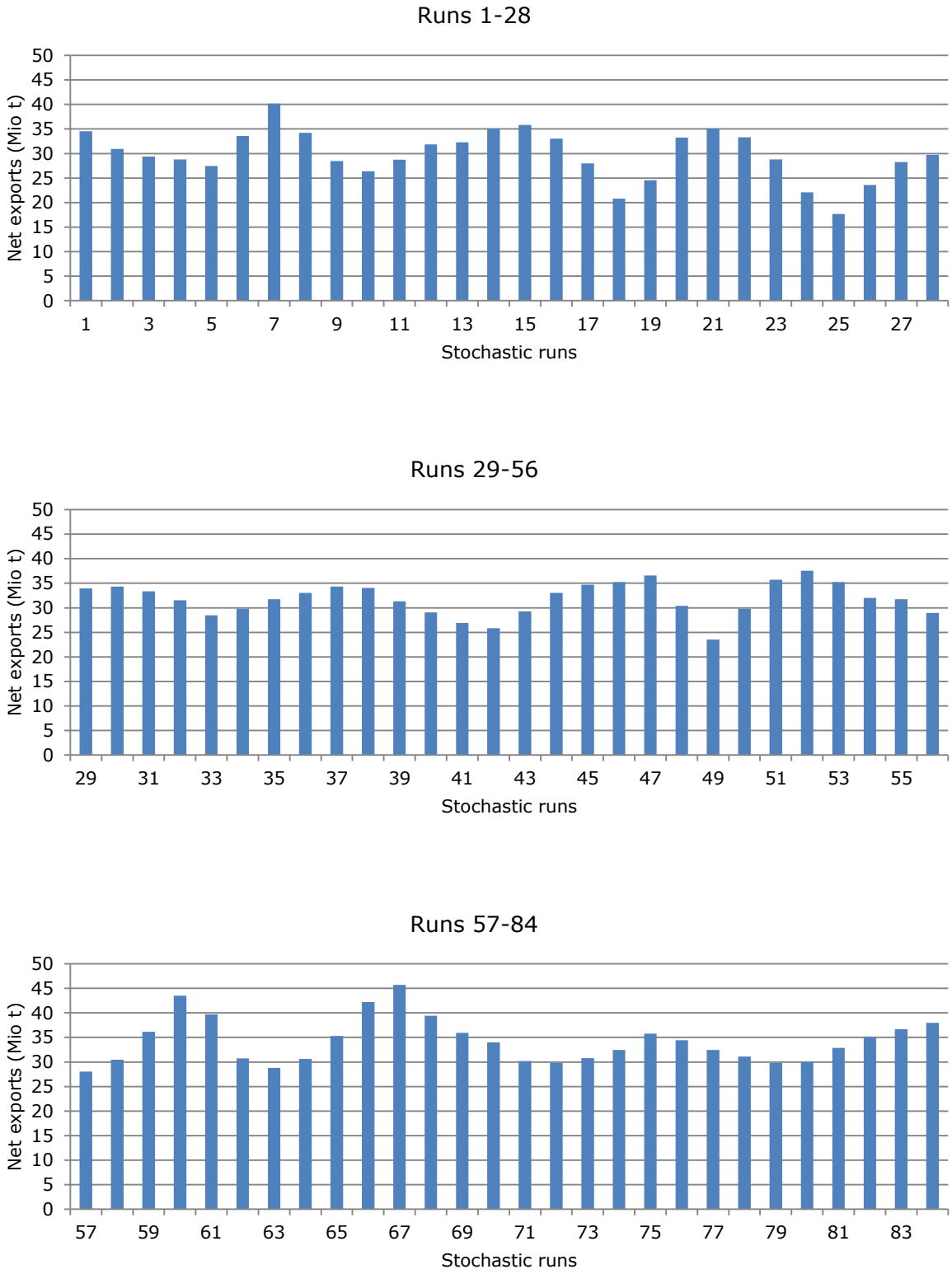
Source: own calculations

Figure 6.6 EU and ROW barley prices obtained in each of the stochastic runs in the baseline scenario



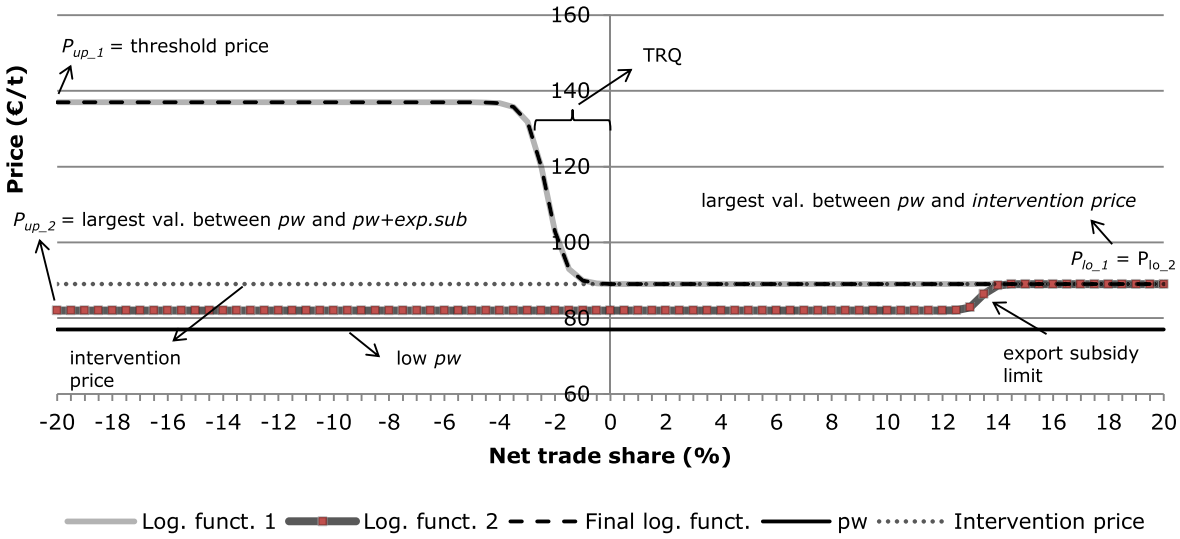
Source: own calculations

Figure 6.7 EU barley net exports obtained in each of the stochastic runs in the baseline scenario



Source: own calculations

Figure 6.8 Price building system for wheat in the EU in the cases when the world market price is low



Note: (Log. funct. 1) and (Log. funct. 2) denote the price logistic functions in ESIM. (Final log. funct.) denotes the final logistic function, which is the largest value between the logistic functions 1 and 2. The logistic function 1 considers TRQs and threshold prices; the logistic function 2 considers export subsidies (exp.sub) and export subsidy limits (pw) denotes the world market price.

Source: own development

Figures 6.4 and 6.6 show that with stochastic yields, the price policies of the EU cereals regime are triggered only in the wheat market. With stochastic yields, the EU wheat net trade position varies between -7.1 and 32.2 million tonnes (see Figure 6.5). In that range of net trade positions, all price policies can be triggered, depending on the level of world market prices. Figure 6.8 shows that with low world market prices and according to each EU net trade position, the corresponding price policy would be triggered. For example, the intervention price is applied in the case of net exports higher than 13.4 million tonnes, which is the export subsidy limit, and the threshold price is applied in the case of imports higher than 3 million tonnes, which is the TRQ. With world market prices below the intervention price (89 €/t), the difference between EU and ROW prices may be large (see e.g. runs 25 and 26 in Figure 6.4). With world market prices above the threshold price, no price policy is simulated, and the EU price equals the world market price.

The price policies simulated for the barley market are not triggered, since the EU net export position in all runs is higher than the export subsidy limit (8.6 million tonnes) (see Figures 6.6 and 6.7). For barley, in net export position above the export subsidy limit, the EU price equals the world market prices, since no intervention is simulated.

The domestic price of the EU for rapeseed is also always at the world market price level since rapeseed is already liberalized in the baseline scenario. Thus, the EV and the CV of the EU are equal to those of the ROW.

Liberalization of the EU cereals regime market results in lower wheat prices in the EU and higher in the ROW. This occurs because the EU removes the policies which maintained the EU price level above the world market price level. As a result, EU production and net exports decreases, and world market prices increase. The EU and the ROW markets for barley also react slightly to the free trade scenario, with a small decrease of prices. This is a result of the cross relationship between the markets of barley and wheat. The reduction in wheat prices motivates an increase in production of barley, which results in a slight decrease in the world market prices of barley. The reaction on the prices of rapeseed is very small, because the cross relationship between rapeseed and wheat is also small and because rapeseed demand is the most elastic (in comparison to wheat and barley), which results in the small price changes observed.

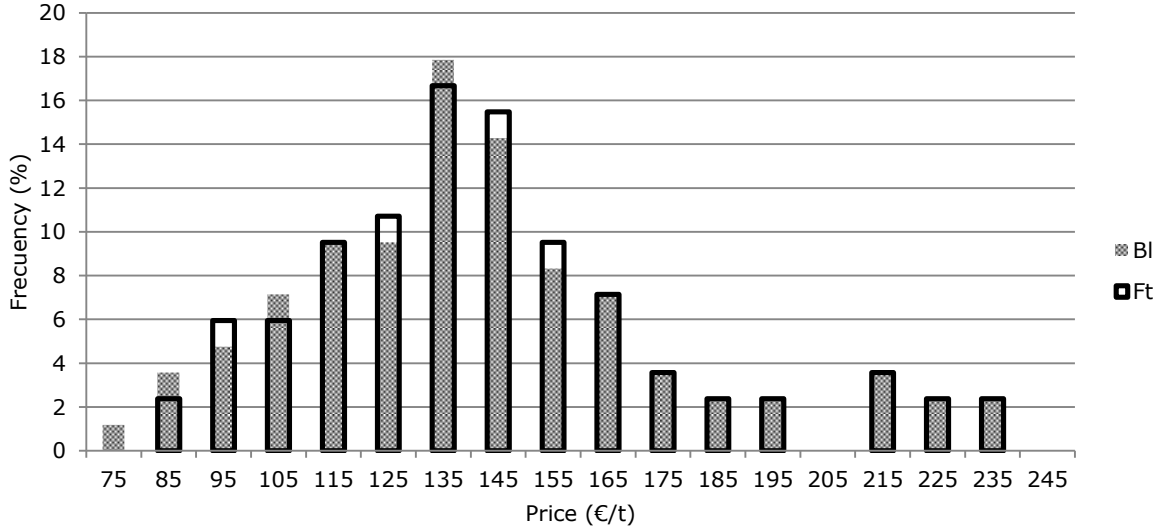
Finally, free trade results in an increase in wheat price uncertainty in the EU (CV of wheat prices increases with free trade) and in a decrease in the ROW. This occurs since with the free trade scenario the EU prices for cereals are allowed to take values below the intervention price. In this way, the reaction of EU supply to world market prices increases, which results in a reduction of the uncertainty of world market prices (the CV of wheat prices in the ROW is reduced).

6.3 Alternative Presentation Forms of Output Variables

Stochastic model results can be presented in many ways, e.g., using the EV and CV as done in Chapters 5 and 6. Alternative presentation forms are box-and-whisker plots (box graphs depicting the quartiles of the results) or histograms. These alternatives present graphically the distribution of model results, which might also be of interest to the analyst. For example, Figure 6.9 shows how the probability distributions of wheat world market prices in the baseline and the free trade scenarios have elongated tails to the right. Also, it can be seen that with free trade, the probability of low world market prices decreases.

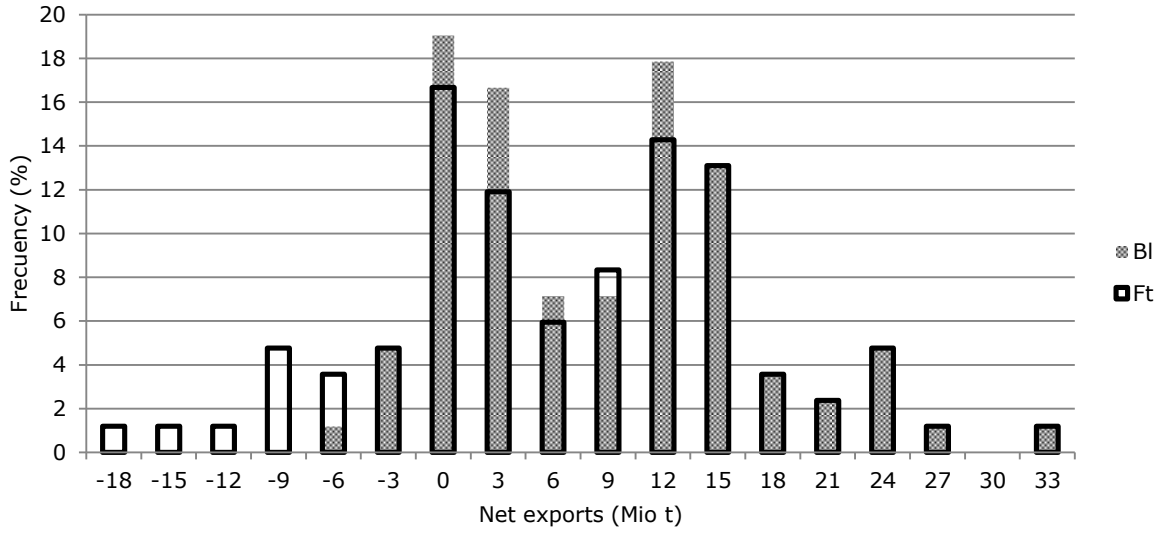
Figure 6.10 shows the histogram of EU wheat net exports obtained in the baseline scenario and under free trade. It can be seen that with free trade the probability distribution of the EU net trade gets wider.

Figure 6.9 Histogram of wheat world market prices in the baseline (Bl) and the free trade (Ft) scenario



Source: own calculations

Figure 6.10 Histogram of the EU wheat net exports in the baseline (Bl) and the free trade (Ft) scenario



Source: own calculations

7 Conclusions and Outlook

7.1 Conclusions

In this study, the accuracies of 8 quadratures obtained with different combinations of the following elements are explored: i) the numerical integration formulas of degree 3 of precision for multivariate standard normal distributions from Artavia et al. (2009) and from Arndt (1996); ii) different arrangements of the stochastic variables (A1 and A2) in the covariance matrix; and iii) different methods to introduce the covariance matrix to the quadratures (via Cholesky decomposition or the diagonalization method). The quadratures are tested in ESIM, which is a large-scale, complex partial equilibrium simulation model of global agricultural markets. The evaluation of the accuracy of the different quadratures is achieved through comparison with an approximated true value of model results.

It was found that in simulation models with isoelastic supply and demand functions which are inelastic, like the stochastic version of ESIM, the use of Artavia et al.'s formula or Arndt's formula is crucial for the determination of the accuracy of the quadratures. With high n, the quadratures based on Artavia et al.'s formula result in samples with a concentration of points near the center and some extreme points at the latter ends of the tails of the marginal probability distributions of the stochastic variables. Since with inelastic supply and demand functions, positive and negative shocks of equal magnitude result in price changes of different sizes, quadratures using Artavia et al.'s formula yield inaccurate results. The quadratures based on Arndt's formula result in sample points which avoid the tails of the marginal distributions of the stochastic variables; however, a systematic concentration of points in one or some parts of the domain of integration cannot be identified. These quadratures yield greater accuracy.

For models or markets with strong asymmetries (e.g. price policies which result in abrupt level changes of output variables at threshold points, large regions that strongly determine the outcome of model results, or isoelastic supply and demand functions that are inelastic) as is the case for wheat in ESIM, the arrangement of the coordinate system or the selected method to induce the covariance matrix may also have a significant effect on the accuracy of the quadratures. In these cases, an analysis of the stability of results is

recommended. This can be done by a repetitive model solves with different arrangement of the stochastic variables in the covariance matrix or by using different methods to introduce correlation.

With the quadratures using Arndt's formula and with markets where the effect of the different regions or variables on model outcomes are more homogenous, as in the case of rapeseed in ESIM, the selection of different arrangements of the stochastic variables in the covariance matrix or of different methods to introduce covariance may not have a significant effect on the accuracy of the quadratures.

The quadratures based on Arndt's formula give an accurate estimate of the uncertainty of model results in ESIM, and simplify stochastic analyses by strongly reducing the number of solves required compared to the LHS sampling method. However, if models are highly asymmetric or with many and strong threshold conditions, the quadratures may lose accuracy. In these cases, the alternative of using Monte Carlo-based approaches should be evaluated. Factors such as higher computational and management costs vs. accuracy gains must be considered.

In addition to the analysis of the accuracy of different quadrature formulas, this study also contributed: i) a summary of the theoretical background of numerical integration formulas, ii) a theoretical analysis of the quadrature formulas from Artavia et al. (2009) and Arndt (1996), iii) a theoretical and practical demonstration on how to induce a desired covariance matrix to standard normal quadratures, iv) a documentation of the stochastic version of ESIM from 2011, and v) an illustration of the relevance of and gain of information from stochastic modeling. The main conclusions are given below.

The summary of the theoretical background on the numerical integration of single and multiple integrals highlighted that interpolation is the basic concept of the approximation techniques and that its accuracy depends on the degree of precision of the quadrature formula, on the smoothness of the integrand and on the capacity of the integrand of being approximated by polynomials.

The theoretical analysis of the formulas from Artavia et al. and Arndt reassure the conclusions about the characteristics of the distribution of points of the quadratures tested. In the analysis, it is shown that both formulas use Stroud's *n*-octahedron, but with different rotations. Furthermore, it is shown that for multivariate standard normal distributions, with the rotation from Arndt's *n*-octahedron, the discrete approximations (samples) of the marginal PDF of the stochastic variables take minimum and maximum values that are given by $\pm\sqrt{2}$. With Artavia et al.'s formula, the vertices of the *n*-octahedron lie on the coordinate axes and the minimum and maximum values of the samples are given by $\pm\sqrt{n}$. Thus, with high n, the samples are composed of zeros and two extreme

points at the latter ends of the tails of the marginal distributions of the stochastic variables. Since the samples are similar before and after the linear transformations of the quadratures, it appears that the rotations of the transformed octahedrons are close to the rotations of the octahedrons before the transformation.

The theoretical and practical demonstration on how to induce a desired covariance matrix to multivariate standard normal quadratures, $\mathbf{\Gamma}$, concludes that the problem is reduced to expressing the desired covariance matrix in the form $\mathbf{A}\mathbf{A}^{\mathrm{T}}$, where $\mathbf{A}$ is a regular square matrix. Once $\mathbf{A}\mathbf{A}^{\mathrm{T}}$ is computed, the linear transformation $\mathbf{A}\mathbf{\Gamma}$ results in the final quadratures with the desired moments.

The application of the stochastic version of ESIM from Chapter 6 shows that multi-market stochastic modeling is a tool which permits the analysis of market uncertainties. For example, it allowed studying the effect of the different price policies on prices with different EU net trade positions and with different world market price levels. Also, it allowed studying the consequences of the liberalization of the EU cereals regime on the EV and CV of commodity prices in the EU and the ROW.

7.2 Outlook

As a future research agenda for the refinement of the application of degree 3 efficient quadrature formulas in large-scale simulation models, one may test rotations from the Q^n slightly different from the one obtained with Arndt's formula. This may result in the inclusion of values from the tails of the marginal distributions of the stochastic variables and in a higher diversity of points. On the other hand, the dependency on the arrangement of the coordinates may increase.

Another area for future research is to test the stability of the quadratures based on Arndt's formula with higher dimensions. Will the samples of the marginal distributions of the stochastic variables always avoid the tail ends of the distributions? Furthermore, one may also test whether and to what extent higher order quadratures approximate the LHS4000 values more accurately.

Finally, the 'efficient quadratures' can be evaluated in other partial equilibrium and general equilibrium models at the same time. Through comparison of the performance of the quadratures and of the model characteristics, factors affecting the accuracy of the quadratures may be identified.

8 References

Anderson, K. and E. Valenzuela (2008), Estimates of Distortions to Agricultural Incentives: 1955 to 2007, spreadsheet at www.worldbank.org/agdistortions, World Bank, Washington D.C.

Arndt, C. (1996), An Introduction to Systematic Sensitivity Analysis via Gaussian Quadrature. GTAP Technical Paper No. 2.

Artavia, M., Deppermann, A., Filler, G., Grethe, H., Häger, A., Kirschke, D., Odening, M. (2010a), Ertrags- und Preisinstabilität auf Agrarmärkten in Deutschland und der EU, Paper presented at the 50th Conference of the GeWiSoLa, September 29 - October 1, Braunschweig, In: http://www.vti.bund.de/fileadmin/dam_uploads/Institute/LR/lr_de/lr_de_gewisola2010/lr_de_beitraege/C2_2.pdf

Artavia, M., Deppermann, A., Filler, G., Grethe, H., Häger, A., Kirschke, D. and M. Odening (2010b), Ertrags- und Preisinstabilität auf Agrarmärkten in Deutschland und der EU – Betriebwirtschaftliche und agrarpolitische Implikationen. In: Rentenbank: Auswirkungen der Finanzkrise und volatile Märkte auf die Agrarwirtschaft. Schriftenreihe der Rentenbank, Band 26: 53-87.

Artavia, M. and D. Kirschke (2010), EU East Enlargement and Instability on CEEC Agricultural Markets, Final Report prepared for the Deutsche Forschungsgemeinschaft (DFG).

Artavia, M., Möller, T., Grethe, H. (2008), Including Stochastic Elements in the Model Analysis, Draft Final Deliverable, Paper presented for the project No. AGRI-2007-G4-12: Study on Modeling Work in the Field of Agriculture – 2, financed by the European Commission.

Artavia, M., Möller, T., Grethe, H. and G. Zimmermann (2009), Correlated Order Three Gaussian Quadratures in Stochastic Simulation Modelling, Selected Paper for the 12th Annual Conference on Global Economic Analysis, 10-12 June, Santiago de Chile. In: https://www.gtap.agecon.purdue.edu/resources/download/4380.pdf

Banse, M., Grethe, H. and S. Nolte (2005), European Simulation Model (ESIM) in GAMS: Model Documentation, Model documentation prepared for DG AGRI, European Commission.

Banse, M., Grethe, H., Nolte, S. and O. Balkhausen (2007), European Simulation Model (ESIM): Draft Model Documentation, Prepared for DG AGRI, European Commission.

DeVusyt, E. A. (1993), Moment Preserving Approximations of Multivariate Probability Distributions for Decision and Policy Analysis: Theory and Applications, Ph.D. Thesis, Purdue University.

European Commission (2011), The EU Cereals Regime, European Commission, Directorate-General for Agriculture and Rural Development, Unit C5.

FAPRI-MU (2011), FAPRI-MU Stochastic U.S. Crop Model Documentation, FAPRI-MU report No. 09-11.

FAOSTAT (2009), Trade domain, sub-domain TradeSTAT, retrieved in November 2009, In: http://faostat.fao.org.

FAOSTAT (2010), Agricultural production domain, retrieved in July 2010, http://faostat.fao.org.

Gujarati, D. N. (2003), Basic Econometrics, Mc Graw Hill, 4th ed.

Hammer, P. C. and A. H. Stroud (1956), "Numerical Integration Over Simplexes", Mathematical Tables and Other Aids to Computation, 10(55): 137-139.

Haber, S. (1970), "Numerical Evaluation of Multiple Integrals", SIAM Review, 12(4): 481-526.

Hertel, T. W., Reimer, J. J., and Valenzuela, E. (2005), "Incorporating commodity stockholding into a general equilibrium model of the global economy", Economic Modelling, 22: 646-664.

Hertel, T. W., W. Martin and A. Leister. (2010), "Potential Implications of a Special Safeguard Mechanism in the World Trade Organization: the Case of Wheat", World Bank Economic Review 24(2):330-359.

Hildebrand, F. B. (1987), Introduction to Numerical Analysis, Dover Publication, 2nd ed., N.Y., USA.

Hinderer, K. (1972), Grundbegriffe der Wahrscheinlichkeitstheorie, Springer-Verlag, Berlin.

Horridge, J. M. and K. Pearson (2011), Systematic Sensitivity Analysis with Respect to Correlated Variations in Parameters and Shocks, GTAP Technical Paper No. 30.

Jechlitschka, K., Kirschke, D. and G. Schwarz (2007), Microeconomics Using Excel, Routledge.

Maxwell, J. C. (1877), "On Approximate Multiple Integration between Limits and Summation", Proc. Cambridge Philos. Soc., 3: 39-47.

Metropolis, N. and S. Ulam (1949), "The Monte Carlo Method", Journal of the American Statistical Association 44(247): 335-341.

Miller, A. C. and T. R. Rice (1983), "Discrete Approximations of Probability Distributions", Management Science, 29(3): 352-362.

OECD (2003), OECD Agricultural Outlook 2003-2008, Paris.

Preckel, P. V. and E. A. DeVuyst (1992), "Efficient Handling of Probability Information for Decision Analysis under Risk", American Journal of Agricultural Economics, 74(3):655-662.

Preckel, P. V., Verma, M., Hertel, T. and W. Martin (2010), Gaussian Quadrature with Correlation and Broader Sampling, Paper prepared for the 14th Annual Conference on Global Economic Analysis, 16-18 June 2011, Venice, Italy.

Schürer, R. (2008), HIntLib Manual, Distributed with version 0.0.13 of HIntLib.

Stoer, J. and R. Bulirsch (2002), Introduction to Numerical Analysis, 3rd ed., Springer-Verlag New York, Inc.

Stroud, A. H. (1957), "Remarks on the Disposition of Points in Numerical Integration Formulas", Mathematical Tables and Other Aids to Computation, 11(60): 257-261.

Stroud, A. H. (1960), "Numerical Integration Formulas of Degree Two", Mathematics of Computation, 14(69): 21-26.

Thacher, H. C. (1957), "Optimum Quadrature Formulas in s Dimensions", Mathematical Tables and Other Aids to Computation, 11(59): 189-194.

Trench, W. (2011), Introduction to Real Analysis, free edition.

Tyler, G. W. (1953), Numerical Integration of Functions of Several Variables, Canadian Journal of Mathematics, 5: 813-821.

Verhoog et al. (2008), Potentials of a Harmonised Database for Agricultural Market Modelling, JRC Scientific and Technical Reports, IPTS, Seville.

Vose, D. (2000), Risk Analysis: A Quantitative Guide, 2nd Ed., John Wiley & Sons Ltd.

Zimmermann, G. (2010), Personal communication.

9 Annex

Annex I: Product Coverage and Activities in ESIM

Product	Farm supply	Processing supply	Human demand	Seed demand	Feed demand	Processing demand
Crops						
Common wheat	X		X	X	X	X
Durum wheat	X		X	X		
Barley	X		X	X	X	
Corn	X		X	X	X	
Rye	X		X	X	X	
Other grains	X		X	X	X	
Rice	X		X	X		
Sugar	X		X			X
Potatoes	X		X	X	X	
Sunflower seed	X		X	X	X	X
Soybeans	X		X	X	X	X
Rapeseed	X			X		X
Manioc	X		X		X	
Fodder	X				X	
Silage maize	X				X	
Animal Products						
Raw milk	X				X	X
Sheep meat	X		X			
Beef	X		X			
Pork	X		X			
Poultry	X		X			
Eggs	X		X			
Processed products						
Sunflower oil		X	X			X
Sunflower cake		X			X	
Soy oil		X	X			
Soy cake		X			X	
Rape oil		X	X			X
Rape cake		X			X	
Palm oil	X		X			X
Cheese		X	X			
Skim milk powder		X	X		X	
Butter		X	X			
Other dairy products		X	X			
Milk		X	X			
Biodiesel		X	X			
Bioethanol		X	X			
Other products						
Pasture	X				X	
Voluntary set aside	X					
Other energy					X	
Other protein					X	
Gluten feed					X	

Source: Banse et al. (2007)

Annex II: Overview of the File Structure of the Deterministic Version of ESIM

File Names	Content
ESIM.GMS	Deciding on scenario or version
SETS.GMS	Definition of static and dynamic sets
ALIAS.GMS	Definition of aliases
PARAMETERS.GMS	Definition of parameters
READING-DATA.GMS	Writing all base data and parameters to the gdx file GSE_XLS.GDX
LOAD-P.GMS	Reading of behavioural parameters and feed rates from PARAMETER.XLS
LOAD-D.GMS	Reading base data from DATA_2000.XLS, DATA_2001.XLS, DATA_2002.XLS, and FINAL_PRICES.XLS Calculation of three year averages
VARIABLES.GMS	Definition of variables
ASSIGNMENTS.GMS	Assignments of base values to variables Assignments of some parameter variables
GSE-READING-DATA.GMS	Unloads parameter saved in the gdx file GSE_XLS.GDX
MACRO-DATA.GMS	Contains values for macroeconomic parameters like exchange rate
CAP.GMS	Contains values for various policy parameters in the base and over the simulation period
CALIBRATION.GMS	
CALIBRATION-DATA.GMS	Definition of subsistence milk share
ADJUST-DAIRY.GMS	Supplementing missing dairy product data Adjusting dairy data such that milk balances are consistent
CHECK-CONSIST.GMS	Calibration of world net exports to zero (adjusting supply in ROW) Calibration of base feed rates
ASSIGN-DYNAMIC-SETS.GMS	Assignment of dynamic sets (which are defined in SETS.GMS)
CALIBRATION-PARA.GMS	Calculation of trade shares Preliminary definition of upper and lower price bounds Calculation of policies for the accession candidates Definition of shadow prices Assignment of finally valid domestic prices from LOGIT-CALIB.GMS Calculation of per hectare premia under SFP and SAPS
LOGIT-CALIB.GMS	Generation of a set of finally valid domestic base prices
CALC-PARA.GMS	Calculation of seed parameters Calculation of crushing rates Calculation of crushing elasticities Calculation of intercepts for all behavioural equations
MAKEGSESHOW.GMS	Functionality only for ESIM in GSE: Defining and tagging input parameters which cannot be edited and writing those to a gdx file SHOWONLY.GDX
MODEL.GMS	Fixing of variables which are zero in the base at zero Definition of equations Solve for the base period Solve over the complete simulation period
CREATERESULTS.GMS	Writing of all solution values of variables and values of various parameters to result parameters
OUTPUT.GMS	Writing all results to a gdx file RESULTS.GDX Tagging of all result parameters
RESULT-TRANSFORM.GMS	Calculation of various results based on core results saved in CREATERESULTS.GMS: aggregation, indices, budgetary implications…

Source: Banse et al. (2005)

Annex III: Overview of Equations in ESIM

Supply equations			
Supply (MCP equation 1, ten specifications)			
(1) Supply of crops in European countries	$SUPPLY_{one,crops}$	=	$EFAREA_{one,crops} \cdot YIELD_{one,crops}$
(2) Supply of energy crops in European countries	$SUPPLY_{i\_biofuel,crops}$	=	$EFAREA_{one,i\_biofuel} \cdot YIELD_{one,i\_biofuel} +$ $EFAREA_{one,i\_biofuel}$ on set aside $YIELD_{one,i\_biofuel}$
(3) Supply of crops in other countries	$SUPPLY_{rest,crops}$	=	$f\ (PP_{rest,crops}, tp\_gr_{rest,crops})$
(4) Supply of animal products	$SUPPLY_{cc,livest}$	=	$f\ (PI_{cc,livest}, FCI_{cc,livest}, lab\_ind_{cc}, cap\_ind_{cc}, int\_ind_{cc}, tp\_gr_{one,livest}, subs\_milk_{cc}, FDEM\_MLK_{cc,"milk"})$
(5) Supply of oilseed products	$SUPPLY_{cc,ospro}$	=	$oilsd\_c_{cc,ospro,oilseed} \cdot PDEM_{cc,oilseed}$
(6) Supply of residual feed	$SUPPLY_{cc,fedres}$	=	$f\ (PD_{cc,fedres}, tp\_gr_{cc,feedres})$
(7) Supply of processed dairy products	$SUPPLY_{cc,dairy}$	=	$MPDEM_{cc,oilseed}/coeff\_d_{dairy}$
(8) Supply of SMP	$SUPPLY_{cc,"smp"}$	=	$coeff\_smp_{cc} \cdot SUPPLY_{cc,"butter"}$
(9) Supply of biofuels	$SUPPLY_{cc,energy}$	=	$f\ (PD_{cc,energ}, BCI_{cc,energ}, pdem\_tr_{cc,energ})$
(10) Supply of gluten feed	$SUPPLY_{cc,"gf"}$	=	$\sum_{i\_ethanol} coef\_p\_bf_{i\_ethanol} PDEM_{cc,i\_ethanol}$
Demand equations			
Human demand (MCP equation 2, one specification)			
(11) Human demand	$HDEM_{cc,comm.}$	=	$f\ (PC_{cc,comm}, pop\_gr_{cc}, inc\_gr_{cc}, subs\_milk_{cc}, hdem\_tr_{cc,comm})$
Seed demand (MCP equation 3, two specifications)			
(12) Seed demand in European countries	$SDEM_{one,crops}$	=	$Seed\_c_{one,crops} \cdot EFAREA_{one,crops}$
(13) Seed demand in other countries	$SDEM_{rest,crops}$	=	$Seed\_c_{rest,crops} \cdot SUPPLY_{rest,crops}$
Processing demand for oilseeds and milk (MCP equation 4, three specifications)			
(14) Processing demand for oilseeds	$PDEM_{cc,oilseed}$	=	$f\ (PD_{cc,oilseed}, PD_{cc,ospro}, pdem\_tr_{cc,oilseed})$
(15) Processing demand for biofuel inputs	$PDEM_{cc,i\_biofuel}$	=	$\sum_{energ} PDEM\_BF_{cc,energy,i\_biofuel}$
(16) Processing demand for milk	$PDEM_{cc,milk}$	=	$\sum_{mlkproc} MPDEM_{cc,mlkproc}$
Processing demand for various dairy products (MCP equation 5, one specification			
(17) PDEM raw milk for single dairy prod.	$MPDEM_{cc,,mlkproc}$	=	$f\ (PD_{cc,"milk"}, PD_{cc,mlkproc})$
Feed demand for all feed products except milk (MCP equation 6, one specification)			
(18) Feed demand	$FDEM_{cc,feed}$	=	$feed\_exog_{cc,feed} + \sum_{livest} FRATE_{cc,feed,livest} \cdot SUPPLY_{cc,livest}$
Feed demand for milk (MCP equation 7, one specification)			
(19) Feed demand for milk	$FDEM\_MLK_{cc,"milk"}$	=	$feed\_milk_{cc} \cdot SUPPLY_{cc,"milk"}$
Feed rates for livestock (MCP equation 8, one specification)			
(20) Feed rate	$FRATE_{cc,feed,livest}$	=	$f\ (PD_{cc,feed}, tp\_fr_{cc,livest})$
Total use (MCP equation 9, one specification)			
(21) Total use	$TUSE_{cc,comm.}$	=	$HDEM_{cc,comm} + SDEM_{cc,comm} + PDEM_{cc,comm} + FDEM_{cc,comm.} + FDEM\_MLK_{cc,comm.}$

Area and Yield Equations

Yield (MCP equation 10, three specifications)

(22) For non-quota crops	$YIELD_{cc,crops}$	=	f ($PP_{cc,crops}$, int_ind_{cc}, lab_ind_{cc}, $tp\_gr_{cc,crops}$)
(23) For non-quota crops	$YIELD_{cc,qu}$	=	f ($PP_{cc,crops}$, $PSH_{cc,qu}$, int_ind_{cc}, lab_ind_{cc}, $tp\_gr_{cc,crops}$)
(24) For energy crops on set aside	$YIELD_{cc,sa\_crp}$	=	$YIELD_{cc,en\_crp}$

Unrestricted area per product (MCP equation 11, three specifications)

(25) Unrestricted product specific area	$ALAREA_{one,crops}$	=	f ($PI_{cc,crops}$, lab_ind_{cc}, int_ind_{cc}, cap_ind_{cc})
(26) For sugar	$ALAREA_{one,crops}$	=	f ($PI_{cc,crops}$, lab_ind_{cc}, int_ind_{cc}, cap_ind_{cc})
(27) For biofuels crops on set aside land	$ALAREA_{one,sa\_crp}$	=	f (PI_{cc,sa_crp}, lab_ind_{cc}, int_ind_{cc}, cap_ind_{cc})

Total unrestricted area (potentially to be scaled) (MCP equation 12, two specifications)

(28) Total unrestricted area	$AREA\_UN_{one}$	=	$\sum_{crops} ALAREA_{cc,crops} - EFAREA_{cc,"sugar"} - EFAREA_{cc,"gras"}$
(29) Total unrestricted area if no sugar quota exists	$AREA\_UN_{one}$	=	$\sum_{crops} ALAREA_{cc,crops} - EFAREA_{cc,"gras"}$

Effective product specific area (MCP equation 13,six specifications)

a) With area scaling to meet base area

(30) Restricted product specific area	$EFAREA_{one,crops}$	=	$ALAREA_{one,crops}/AREA\_UN_{one} \cdot ((area\_to_{one} - EFAREA_{one,"sugar"} - EFAREA_{one,"gras"} - OBLSETAS_{one}) \cdot area\_gr_{one})$
(31) Area for non-scaled products (incl. biofuel products on vol. set aside)	$EFAREA_{one,qu\ (or\ "gras")}$	=	$ALAREA_{one,qu\ (or\ "gras")}$
(32) Restricted product specific area for countries without sugar quota	$EFAREA_{one,crops}$	=	$ALAREA_{one,crops}/AREA\_UN_{one} \cdot ((area\_to_{one} - EFAREA_{one,"gras"} - OBLSETAS_{one}) \bullet area\_gr_{one})$
(33) Area limit for energy crops on set aside (applies under (a) and (b))	$EFAREA_{one,i\_biofuels}$	=	$Min\ (ALAREA_{one,i\_biofuels}, ALAREA_{one,i\_biofuels} / \sum_{i\_biofuels} ALAREA_{one,i\_biofuels} \bullet (EFAREA_{one,"setaside"} + OBLSETAS_{one} \bullet area\_gr_{one})$

b) Without scaling, but with maximum area restriction

(34) Restricted product specific area	$EFAREA_{one,crops}$	=	$Min\ (ALAREA_{one,crops} + ALAREA_{one,crops}/AREA\_UN_{one} \bullet marg\_land_{one} \bullet chg\_oblsetas_{one}, ALAREA_{one,crops}/AREA\_UN_{one} \cdot ((area\_to_{one} - EFAREA_{one,"sugar"} - EFAREA_{one,"gras"} - OBLSETAS_{one}) \cdot area\_gr_{one})$
(35) Restricted product specific area for countries without sugar quota	$EFAREA_{one,crops}$	=	$Min\ (ALAREA_{one,crops} + ALAREA_{one,crops}/AREA\_UN_{one} \bullet marg\_land_{one} \bullet chg\_oblsetas_{one}, ALAREA_{one,crops}/AREA\_UN_{one} \cdot ((area\_to_{one} - EFAREA_{one,"gras"} - OBLSETAS_{one}) \cdot area\_gr_{one})$

Effective obligatory set aside (MCP equation 14, one specification)

(36) Effective obligatory set aside area	$OBLSETAS_{one}$	=	$area\_gc_{one} \cdot setaside_{one} \cdot (1 - smlprod_{one}) \cdot (1 - slippage_{one})$

Grand cultures area (MCP equation 15, one specification)

(37) Grand cultures area	$EFAREA\_GC_{one}$	=	$\sum_{comcrop} EFAREA_{one,comcrop} + OBLSETAS_{one}$

Direct Payment Equations

Direct payments (MCP equation 16, two specifications)

(38) Direct payments for crops per ton	$DIRPAY_{one,crop}$	=	$DIRP\_t\_nc_{one,crop} / YIELD_{one,crop} \bullet YIELD0_{one,crop}$
(39) Direct payments for livest. per ton	$DIRPAY_{one,livest}$	=	$DIRP\_t\_nc_{one,livest}$

Price equations

Lower bound of Logit-function (MCP equation 17, four specifications)

(40) For NMS prior to accession	$P\_LO_{one,it}$	=	$PW_{it}/exrate_{one} \cdot (1 + subs\_ad_{one,it})$
(41) For EU and products with interv. Price	$P\_LO_{one,it}$	=	$MAX(PW_{it}/exrate_{one}, intpr_{one,it} + exstab_{one,it})$
(42) For EU and delayed integration	$P\_LO_{delay\_r,delay\_c}$	=	$delay_{delay\_r,delay\_c} PD_{"EU15",delay\_c} exrate_{euro}/exrate_{one}$
(43) For EU products without interv. Price	$P\_LO_{one,it.}$	=	$PW_{it}/exrate_{one}$

Upper bound of Logit-function (MCP equation 18, three specifications)

(44) For NMS prior to accession	$P\_UP_{one,it}$	=	$PW_{it}/exrate_{one} \ (1+tar\_ad_{one,it})$
(45) For EU products with thresh. price	$P\_UP_{one,thresh}$	=	$MAX(PW_{thresh}/exrate_{one}, thrpr_{one,thresh})$
(46) For EU products without thresh. Price	$P\_UP_{one,it}$	=	$PW_{it}/exrate_{one} \cdot (1+tar\_ad_{one,it}) + sp\_d_{one,it}$

Second upper bound of logit function because of export subsidy (MCP equation 21, one specification)

(47) Second upper bound	$P\_UP_{one,it}$	=	$MAX(exp\_sub_{one,it}, qual\_ad_{one,it}) + PW_{it}/exrate_{one}$

Price transmission function (MCP equation 19, five specifications)

(48) World market price	$PD_{row,it}$	=	PW_{it}
(49) PW transmission for EU markets without export subsidies	$PD_{one,it}$	=	Logistic function
(50) PW transmission for EU markets with export subsidies	$PD_{one,it}$	=	Logistic function
(51) PW transmission for new members with delayed integration	$PD_{one,it}$	=	Logistic function
(52) PW transmission for non-members	$PD_{one,it}$	=	Logistic function

Shadow price determination (MCP equation 20, three specifications)

(53) Shadow price livestock	$PSH_{one,livest}$	=	f ($quota_{one,livest}$, $PI_{one,livest}$, $FCI_{one,livest}$, lab_ind_{one}, cap_ind_{one}, int_ind_{one}, $tp\_gr_{one,livest}$)
(54) Shadow price crops	$PSH_{one,crops}$	=	f ($quota_{one,crops}$, $PI_{one,crops}$, lab_ind_{one}, cap_ind_{one}, int_ind_{one}, $tp\_gr_{one,crops}$)
(55) Shadow price voluntary set aside	$PSH_{one,"setaside"}$	=	f ($quota_{one,"setaside"}$, $PI_{one,"setaside"}$, lab_ind_{one}, cap_ind_{one}, int_ind_{one}, $tp\_gr_{one,crops}$)

Wholesale/producer price transmission (MCP equation 21, four specifications)

(56) Producer price if margin	$PP_{cc,nq}$	=	$PD_{cc,nq}/margin0_{one,nq}$
(57) Producer price if no margin	$PP_{cc,nq}$	=	$PD_{cc,nq}$
(58) Incentive price for quota products	$PP_{one,qu}$	=	$MIN(PD_{one,qu}/margin0_{one,qu}, PSH_{one,qu})$
(59) Producer price for energy crops on set aside	PP_{one,sa_crp}	=	PP_{one,en_crp}

Consumer price (MCP equation 22, two specifications)

(60) For agri-food commodities	$PC_{cc,comm}$	=	$PD_{cc,comm} \cdot + pc\_tax_{cc,comm.}$
(61) For mineral oil	$PC_{cc,"cr\_oil"}$	=	$p0_{cc,"cr\_oil"} \cdot + pc\_tax_{cc,"cr\_oil"}$
Determination of producer incentive price (MCP equation 23, four specifications)			
(62) For non quota products	$PI_{cc,nq}$	=	$PP_{cc,nq} + prod\_eff_{cc,nq} \bullet DIRPAY_{cc,nq}$
(63) For quota products	$PI_{cc,qu}$	=	$MIN(PP_{cc,qu} + prod\_eff_{cc,nq} \bullet DIRPAY_{cc,qu}, PSH_{cc,qu})$
(64) For energy crops in EU on non set-aside	$PI_{cc,energ}$	=	$pay\_biof/\sum_{energ} \sum_{member} SUPPLY_{c,energ} \bullet exrate_{"EURO"}/exrate(cc)$
(65) For energy crops in non EU	$PI_{cc,energ}$	=	$PD_{cc,energ}$
Net Price for biofuel inputs (MCP equation 24, two specification)			
(66) Net Price for biodiesel inputs	$NetPD_{cc,i\_biodiesel}$	=	$PD_{cc,i\_biodiesel} - PD_{cc,i\_biodiesel,byproduct}$
(67) Net Price for bioethanol inputs	$NetPD_{cc,i\_bioethnaol}$	=	$PD_{cc,i\_bioethanol} - PD_{cc,i\_bioethanol,byproduct}$
Feed cost index (MCP equation 25, one specification)			
(68) Feed cost index	$FCI_{cc,livest}$	=	$\sum_{feed} FRATE_{cc,feed,livest} \cdot PF_{cc,feed}/ FC\_0_{cc,livest}$
Biofuel input cost index (MCP equation 26, one specification)			
(69) Biofuel input cost index	$BCI_{cc,energ}$	=	$\sum_{i\_biofuel} QUANCES0_{cc,i\_biofuel} \cdot NetPD_{cc,i\_biofuel}/ \sum_{i\_biofuel} QUANCES0_{cc,i\_biofuel}/BCI0$

Equations for CES technology in biofuel production

Determination of unscaled relative input quantities (MCP equation 27, two specifications)			
(70) Biodiesel	$QUANCES_{cc,energ,i\_biodiesel}$	=	CES function
(71) Bioethanol	$QUANCES_{cc,energ,i\_bioethanol}$	=	CES function
Scaling of relative input quantities to add up to supply of biofuels (MCP equation 28, one specification)			
(72) Scaling of input quantities	$QUANCES_{cc,energ,i\_biofuel} / \sum_{QUANCES} QUANCES_{cc,energ,i\_biofuel}$	=	$PDEM\_BF_{cc,energ,i\_biofuel}/convbf_{energy,i\_biofuel}/ SUPPLY_{cc,energ}, T, T$

Other equations

Net exports (MCP equation 29, one specification)			
(73) Net exports	$NETEXP_{cc,it}$	=	$SUPPLY_{cc,it} - TUSE_{cc,it}$
Determination of trade shares (MCP equations 30-32)			
(74) Share of net exports in domestic market volume (EU)	$TRADESHR_{eu,it}$	=	$\sum_{member} NETEXP_{member,it}/\sum_{member} TUSE_{member,it}$
(75) Share of net exports in domestic market volume (delayed integration region)	$TRADESHR_{delay\_r,delay\_c}$	=	$\sum_{delay\_c} NETEXP_{delay\_r,delay\_c}/\sum_{delay\_c} TUSE_{delay\_r,delay\_c}$
(76) Share of net exports in domestic market volume (individual countries)	$TRADESHR_{one,it}$	=	$NETEXP_{one,it}/MAX(SUPPLY_{one,it}, TUSE_{one,it})$
Determination of quality shares (MCP equation 33, one specification)			
(77) Share of high quality exports in domestic market volume	$QUALSHR_{eu,it}$	=	$QUALQUANT_{"eu",it}/\sum_{member} TUSE_{member,it}$
Determination of export subsidy shares (MCP equation 34, one specification)			
(78) Share of export subsidy limit in domestic market volume	$SUBSHR_{eu,it}$	=	$SUBQUANT_{"eu",it}/ \sum_{member} TUSE_{member,it}$
Determination of TRQ shares (MCP equation 35, one specification)			
(79) Share of TRQ in domestic market volume	$TRQSHR_{eu,it}$	=	$TRQ_{"eu",it}/ \sum_{member} TUSE_{member,it}$
World market clearing (MCP equation 36, one specification)			

(80) World market clearing condition	0	=	$\sum_{cc} NETEXP_{cc,it}$
Domestic market clearing (MCP equation 37, one specification)			
(81) Domestic market clearing condition for non tradables	$SUPPLY_{cc,nt}$	=	$TUSE_{cc,nt}$

[a] In the MCP formulation, equations often include a set of different specifications. For example, the specification of the supply equation for animal products in the EU is different from the specification of the EU supply equation for crops. For better readability, all specifications of equations are numbered consecutively in this table and referred to as equations throughout the documentation. In addition, equations are displayed throughout this documentation with the dependent variable at the left hand side – in contrast to the MCP formulation in the GAMS code, in which equations are written as implicit functions such that a zero appears at the left hand side.

Source: Banse et al. (2005)